EXTERNAL SCIENTIFIC REPORT

APPROVED: 26 January 2021

doi:10.2903/sp.efsa.2021.EN-6476

Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0

Emilio Benfenati, Alessandra Roncaglioni, Edoardo Carnesecchi, Matilda Mazzucotelli, Marco Marzo, Andrey Toropov, Alla Toropova, Rossella Baldin, Andrea Ciacci, Simona Kovarich, Luca Sartori, Chihae Yang, Tomasz Magdziarz, Bryan Hobocienski, Aleksandra Mostrag

Istituto di Ricerche Farmacologiche Mario Negri IRCCS, S-IN Soluzioni Informatiche, Molecular Networks GmbH

Abstract

The present document represents the summary of the activities undertaken during the second year of the framework contract (OC/EFSA/SCER/2018/01) with the overall objectives of maintaining, updating and further developing the OpenFoodTox database ("OpenFodTox 2.0"). OpenFoodTox has been developed over the last 8 years to map hazard data as published in EFSA documents (opinions, statements, and conclusions) dealing with chemicals risk assessment of food and feed. The repository holds summary data on identification of chemicals, document descriptors, hazard identification, and hazard characterisation. Overall, OpenFoodTox 2.0 now includes hazard data from more than from 10300 assessments, 2140 EFSA outputs and 5100 chemicals with new entries for 138 new substances, 437 hazard assessments from 181 EFSA documents. The data collection for new data types has been continued with further expansion of the data model to incorporate new data types including physicochemical properties (OHT 1 to 23-5), degradation and bioaccumulation data (OHT 32 and 33), toxicokinetic data (OHT 58), intermediate effects (OHT 201), "New Approach methodologies (NAMs)" and Exposure Information (OHT 301 to 306). OpenFoodTox content was enriched with physico-chemical properties and PK/TK data including quantitative values for key parameters (Cmax, AUC, half-life) obtained from EFSA documents published between 2018 and 2020 and progressively integrated in the database. Overall, 3658 experimental physico-chemical data records for 522 substances (138 EFSA outputs) and 1986 PK/TK records were added for 315 substances (108 EFSA outputs). Case studies on OHT 201 to report mechanistic data are illustrated. Furthermore, preliminary results on new QSAR models are here presented as part of the design of an *in silico* integrative tool allowing for the description and prediction of hazard properties of chemicals for "OpenFoodTox 2".

Key words: EFSA's OpenFooxTox, OECD templates, data model, critical study, hazard assessment, QSAR models.

Question number: EFSA-Q-2020-00159

Correspondence: sc.secretariat@efsa.europa.eu

Acknowledgements: The authors would like to thank Mr Jean-Lou Dorne (EFSA), Ms Jane Richardson (EFSA), and Ms Anastasia Livaniou (EFSA) for valuable support and input provided during the project execution. The authors are grateful to JRC colleagues (Clemens Wittwehr and Alicia Paini) for the valuable discussions on OECD Harmonised Templates (OHTs).

Suggested citation: Emilio Benfenati, Alessandra Roncaglioni, Edoardo Carnesecchi, Matilda Mazzucotelli, Marco Marzo, Andrey Toropov, Alla Toropova, Rossella Baldin, Andrea Ciacci, Simona Kovarich, Luca Sartori, Chihae Yang, Tomasz Magdziarz, Bryan Hobocienski and Aleksandra Mostrag, 2021. Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFSA supporting publication 2021:EN-6476. 46 pp. doi:10.2903/sp.efsa.2021.EN-6476

ISSN: 2397-8325

Summary

The EFSA's Chemical Hazards Database 'OpenFoodTox' aims at mapping hazards data published in outputs from the European Food Safety Authority (EFSA) (opinions, statements, conclusions) with regards to risk assessments of chemicals in food and feed. The database covers the work of many units and panels, 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), PPR panel and PRAPeR unit (Plant Protection Products and their Residues), NDA (Dietetic Products, Nutrition and Allergies). OpenFoodTox was initially developed nine years ago to summarise hazards data for chemicals assessed by EFSA since its creation.

The entire database includes data extracted from the screening of more than 2140 documents (opinions, statements, conclusions) published by EFSA from 2003 to July 2020, holding more than 10000 assessments for about 5100 chemicals (each assessment potentially including multiple (eco-)toxicity endpoints and/or hazard/risk characterisation data). The assessments are classified based on well-defined categories (e.g. pesticides, flavourings, sensory additives, nutrient sources). For example, the repository currently stores: 6013 assessments related to flavourings (2123 substances) as collected from the screening of 317 EFSA documents (opinions and statements); 1560 assessments related to pesticides (1266 substances) as collected from the screening of 554 EFSA documents (conclusions); 406 assessments related to food additives (332 substances) as collected from the screening of 223 EFSA documents (opinions and statements). In addition, following a request from EFSA, dietary reference values (DRV) for nutrients (minerals and vitamins) have been entered within the OFT database to harmonise OFT with the EFSA DRV finder. [1] The DRV values included AI (Adequate Intake), AR (Average Requirement), PRI (Population Reference Intake) and UL (Upper Limit) values. It is acknowledged that most of the UL values have been extracted from SCF (EC Scientific Committee on Food) documents published before the establishment of EFSA. For this reason, the OFT database now includes data published external to EFSA (i.e., SCF documents) and before 2003.

The database can be downloaded from EFSA knowledge junction under: https://doi.org/10.5281/zenodo.780543 and can be consulted from a MicroStrategy Dashboard: https://www.efsa.europa.eu/en/microstrategy/openfoodtox.

At this stage of the project, no major modifications of the standard operating procedures (SOPs) in use for hazard data entry (Appendix A) have been performed. A minor revision to the SOPs was applied and consisted of the assignment of PARAM codes to all chemicals (substances / components) included in the OpenFoodTox 2.0 database. An updated data model design was necessary to map new properties to the database (physicochemical properties and toxicokinetic data) and inclusion of these new types of data has already started. Preliminary results on the case study on OHT 201 on "intermediate effect" available via the IUCLID 6 software tool are provided for i) in vitro skin sensitization data (human Cell Line Activation Test) for dihydroeugenol and ii) ER-alpha agonism and antagonism, TR-beta agonism and antagonism, and binding/docking literature data for fipronil.

Finally, ten different QSARs models have been developed using OpenFoodTox for relevant toxicological endpoints in species of ecological relevance such as the earthworm, Daphnia spp. bobwhite quail, rainbow trout and honey bees. In addition, human toxicology data were collected to build a model for inhalation toxicity.

[1] https://www.efsa.europa.eu/en/interactive-pages/drvs?lang=en

Table of contents

1. Introduction

1.1. Background and Terms of Reference as provided by the requestor

The European Food Safety Authority (EFSA) is the keystone of European Union (EU) risk assessment regarding food and feed safety. In close collaboration with national authorities and in open consultation with its stakeholders, EFSA provides independent scientific advice and clear communication on existing and emerging risks.

EFSA was established in January 2002, following a series of food crises in the late 1990s, as an independent source of scientific advice and communication on risks associated with the food chain.

EFSA was created as part of a comprehensive programme to improve EU food safety, ensure a high level of consumer protection and restore and maintain confidence in the EU food supply.

In the European food safety system, risk assessment is done independently from risk management. As the risk assessor, EFSA produces scientific opinions and advice to provide a sound foundation for European policies and legislation and to support the European Commission, European Parliament and EU Member States in taking effective and timely risk management decisions.

EFSA's remit covers food and feed safety, nutrition, animal health and welfare, plant protection and plant health. In carrying out its work, EFSA also considers the possible impact of the food chain on the biodiversity of plant and animal habitats. The Authority performs environmental risk assessments of genetically modified crops, pesticides, feed additives, and plant pests. In all these fields, EFSA's most critical commitment is to provide objective and independent, science-based advice and clear communication grounded in the most up-to-date scientific information and knowledge.

Since its creation in 2002, EFSA has produced risk assessments for more than 4,900 unique substances in over 1,900 Scientific Opinions, Statements and Conclusions through the work of its scientific Panels, Units, and Scientific Committee.

The EFSA Strategy 2020 includes the strategic objective to "Widen EFSA's evidence base and optimise access to its data". Under this objective, EFSA plans to "migrate towards structured scientific data" as a move towards efficiency, innovation and new methods in risk assessment through structuring of data from monitoring schemes, regulated product applications and EFSA outputs, in agreed formats and based, where possible, on existing international standards, enabling their re-use. Therefore, the EFSA "Information Management Programme" has several projects designed to increase the use of standardised data.

In this context, the EFSA SCER unit has been developing EFSA's Chemical Hazards database "OpenFoodTox" since 2011 in collaboration with the assessment methodology and the evidence management units. OpenFoodTox provides summary toxicological data used by EFSA for setting safe levels (reference points and reference values) of food and feed chemicals in humans, animals and the environment since its foundation in 2002.

The design of OpenFoodTox has taken into account structure templates, namely the OECD harmonised templates (OHTs), for the coherent and harmonised reporting of toxicological data. In addition, the database enables sharing of hazard data with sister agencies and international scientific advisory bodies through EFSA's data warehouse and knowledge junction and the OECD's Global Portal to Information on Chemical Substances (eChemPortal). Within EFSA's strategy 2016-2020, OpenFoodTox 1.0 has been published as part of Open EFSA to give all stakeholders access to all hazard data through: i) the data warehouse in the form a microstrategy tool; ii) publication on the full datasets on EFSA knowledge junction for full download; iii) access through the OECD e-chem portal.

Over the last few months, CEFIC integrated OpenFoodTox with ECHA's REACH database and the QSAR platform VEGA in "AMBIT2": an open source predictive tool providing support to industry, academia, risk assessment bodies and other stakeholders in the chemical risk assessment area.

The overall objective of this procurement is to maintain, update and further develop EFSA's OpenFoodTox 2.0 as an open source tool for EFSA, stakeholders and the entire risk assessment community. In particular, the aim of the 4-year project "Further development and update of EFSA's Chemical Hazards Database OC/EFSA/SCER/2018/01" is to update and further develop EFSA's Chemical Hazards Database (i.e., OpenFoodTox) by accomplishing the three main objectives (or tasks) of the assignment:

- Call objective 1: Data collection to maintain, update, and further develop OpenFoodTox: OpenFoodTox 2.0.
- Call objective 2: Design and update new OHTs for implementing the use of overarching guidance documents in chemical risk assessment and OpenFoodTox.
- Call objective 3: Design of an *in silico* integrative tool allowing description and prediction of hazard properties of chemicals for OpenFoodTox 2.0.

This contract was awarded by EFSA to: Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Molecular Networks GmbH. S-IN, Soluzioni Informatiche srl. Contractor: Istituto di Ricerche Farmacologiche Mario Negri IRCCS, VAT registration number: IT 03254210150.

Contract: 'Further development and update of EFSA's Chemical Hazards Database'

Contract: OC/EFSA/SCER/2018/01

1.2. EFSA's Chemical Hazards Database: OpenFoodTox

As described in the external reports published by EFSA (EFSA, 2014; EFSA, 2015; EFSA, 2017; S-IN, 2018; Benfenati, 2020), EFSA's Chemical Hazards Database "OpenFoodTox" aims at mapping the hazard data included in the documents (opinions, statements, conclusions) on risk assessment in food and feed published by EFSA. The database covers the work of many units and panels, 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). EFSA's Hazards Database stores summary data on the chemical entity, the document details, the hazard identification, and the hazard characterisation/risk characterisation. The latest update and maintenance of the database involved extraction, collection and collation of relevant data included in the EFSA documents that were adopted (and then published) by the Scientific Panels until May 2019 (Benfenati et al., 2020).

Notably, in parallel to data collection and data entry, the SCER Unit is collaborating with the SAS and ITS Units, as well as with the Communications Directorate for the development of a web interface for both internal and external consultation of EFSA's Chemical Hazards Database.

As mentioned above, the data model of EFSA's Chemical Hazards Database, designed by EFSA and finalized by S-IN Soluzioni Informatiche in the context of a previous procurement ("Data collection and data entry for EFSA's Chemical Hazards Database" - NP/EFSA/EMRISK/2011/01), is structured to map the intrinsic properties of the hazard data extracted from EFSA opinions. The data model as implemented in the previous assignments includes a number of modules describing:

- 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 has 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 used to

derive the health-based guidance value (TDI, 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).

The entire database includes data as extracted from the screening of more than 2140 documents (opinions, statements, conclusions) published from 2003 to July 2020, holding more than 10000 assessments for about 5100 chemicals. The assessments are classified based on well-defined categories (e.g. pesticides, flavourings, sensory additives, nutrient sources). For example, the repository currently stores: 6013 assessments related to flavourings (2123 substances) as collected from the screening of 317 EFSA documents (opinions and statements); 1560 assessments related to pesticides (1266 substances) as collected from the screening of 554 EFSA documents (conclusions); 406 assessments related to food additives (332 substances) as collected from the screening of 223 EFSA documents (opinions and statements). In addition, following a request from EFSA, dietary reference values (DRV) for nutrients (minerals and vitamins) have been entered within the OFT database to harmonise OFT with the EFSA DRV finder. [2] The DRV values included AI (Adequate Intake), AR (Average Requirement), PRI (Population Reference Intake) and UL (Upper Limit) values. It is acknowledged that most of UL values have been extracted from SCF (EC Scientific Committee on Food) documents published before the establishment of EFSA. For this reason, the OFT database now includes data published external to EFSA (i.e., SCF documents) and before 2003.

1.3. Interpretation of the Terms of Reference

The overall objective of the underlying project 'Further development and update of EFSA's Chemical Hazards Database' was the maintenance of OpenFoodTox by collecting **all hazard data included in the documents related to the following panels:** 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), PPR panel and PRAPeR unit (Plant Protection Products and their Residues), NDA (Dietetic Products, Nutrition and Allergies). Substances which do not fall within the category of chemicals (e.g., microorganisms, fungi, nematodes, and enzymes) are excluded from OpenFoodTox. The project entailed three activities:

1. Collecting data in order to maintain, update and further develop OpenFoodTox: "OpenFoodTox 2.0".
2. Piloting and updating the design of a new OHT for implementing the use of overarching guidance documents in chemical risk assessment and OpenFoodTox.
3. Designing an *in silico* integrative tool allowing for the description and prediction of hazard properties of chemicals for OpenFoodTox 2.0.

The key activities to meet **objective 1)** can be summarised according to three main subtasks:

- Maintaining and updating all data currently present in OpenFoodTox.
  Data collection and entry of all hazard data assessed by EFSA scientific panels (e.g., ANS, CEF, CONTAM, FEEDAP, PPR and Pesticide Unit, NDA) from the latest dataset published in OpenFoodTox, i.e. May 2019 (Benfenati et al., 2020) using the existing and already published data model. The mapping of EFSA documents (opinions, conclusions) include the following key aspects: i) details of the EFSA document, ii) chemical characterisation (e.g., IUPAC, CAS and EC numbers, SMILES, PARAM, etc..) and composition, iii) hazard data (human, animal and ecotoxicology including genotoxicity, reference points, reference values, target organ toxicity,

[2] https://www.efsa.europa.eu/en/interactive-pages/drvs?lang=en

uncertainty factors, etc.). A frequent transfer of collected data (in XML format) on EFSA's Data Collection Framework (DCF) is performed in order to ensure that the dataset is compliant with EFSA IT standards and to enable future sharing of the data with EU Member States, other EU agencies (ECHA, EMA), international bodies (OECD, FAO/WHO) and third parties (US-EPA).

- Data collection of new content types of physicochemical properties, PK/TK (year 1-2), and exposure studies (year 3).
- Modification of the OpenFoodTox data model necessary for the extension of the database to new substance properties to comply with the OECD Harmonised Templates (OHT). Specifically, with OHT for physicochemical properties (OHT 1 to 23-5), and toxicokinetic data (OHT 58) during year 1-2, and use and exposure information (OHT 301 to 306) (year 3). Moreover, the incorporation of other templates was investigated, including the ones for degradation and bioaccumulation (OHT 32 and 33), and intermediate effects (OHT 201).

The activities within **objective 2)** include:

- Investigating the relevance of existing OHTs and identify potential needs to update and/or extend them in order to facilitate the data interoperability and straightforward exchange of information between various agencies and platforms. To this end, case studies were conducted using OHT 201 and "intermediate effects" data for two tested substances, dihydroeugenol and fipronil.
- Representing the way to achieve the ambitious target of designing a framework to integrate the results of different OHTs for the implementation of EFSA overarching guidance documents on:
    - The use of the weight of evidence (WoE) approach in scientific assessments: https://www.efsa.europa.eu/en/efsajournal/pub/4971;
    - Assessment of the biological relevance of data in scientific assessment: https://www.efsa.europa.eu/en/efsajournal/pub/4970;
    - Uncertainty analysis in Scientific Assessments: https://www.efsa.europa.eu/en/efsajournal/pub/5123;

A preliminary exercise has been conducted considering conazoles, also with reference to the OHTs using the key features and methods in the three guidance documents (e.g. methods as picklists, reporting tables etc.). The feasibility has been evaluated first by screening a large set of conazoles, and then focussing the evaluation on a few representative substances. The WoE guidance will provide the key pillars for the conceptual scheme on how to integrate heterogeneous lines of evidence. These pillars are the relevance, reliability and concordance between multiple values. The EFSA guidance provides examples, also relative to non-testing methods. Similarly, the biological relevance of adverse/beneficial health effects from experimental animal and human studies will be addressed also considering the related EFSA document. The inclusion of data derived from *in vivo*, *in vitro* and *in silico*, requires taking into account the related uncertainty. In addition, uncertainties arising when assessing biological relevance should be addressed and described as part of the WoE assessment. The third document, on uncertainty, is of common interest for both the two previous documents.

The activities regarding **objective 3)** include:
OpenFoodTox 2.0 will provide an integrative tool allowing for the description and prediction of hazard properties of chemicals. For the *in silico* component, models in VEGA represent an advanced system, which has been already applied to OpenFoodToox 1.0 in an initial phase. Beyond simply applying VEGA to fill data gaps, it has been demonstrated that new models can be developed using the experimental data present in OpenFoodTox. The present project will proceed in a comprehensive way, thus developing and applying tools for read across, based on ToxRead. Furthermore, the pharmacokinetic properties of the substances will be used to increase the reliability of the assessment on the substance to be evaluated, moving towards a more robust risk assessment. The novelty of this project is to integrate all these multiple perspectives. The pharmacokinetic perspective will provide a closer view of the different behaviour depending on properties which are not usually addressed with classical *in silico* models. Indeed, our new project will address internal exposure and behaviour, also considering different

behaviour in different species. In this way, a much better modelling ability for the internal exposure can be achieved. We will also address other properties very important for exposure, such as bioconcentration. Joining the exposure modelling with the hazards, better risk assessment can be achieved.

During the first year the specific achievements driving to the final objectives were the initial development of QSAR models based on the OpenFoodTox data to obtain QSAR models more relevant to the chemical space targeted by EFSA activities. These exercises will serve as a basis to choose the best option to integrate the results of the predictions within the OpenFoodTox database.

2. Methodologies

2.1 Task 1/Objective 1 - Data collection for OpenFoodTox 2.0

2.1.1 Data entry

2.1.1.1 Hazard Data

A suitable ad hoc accessory IT platform (developed in-house) aids and supports the activity workflow entailed in the present assignment: data collection, entry and submission (see Figure 1).

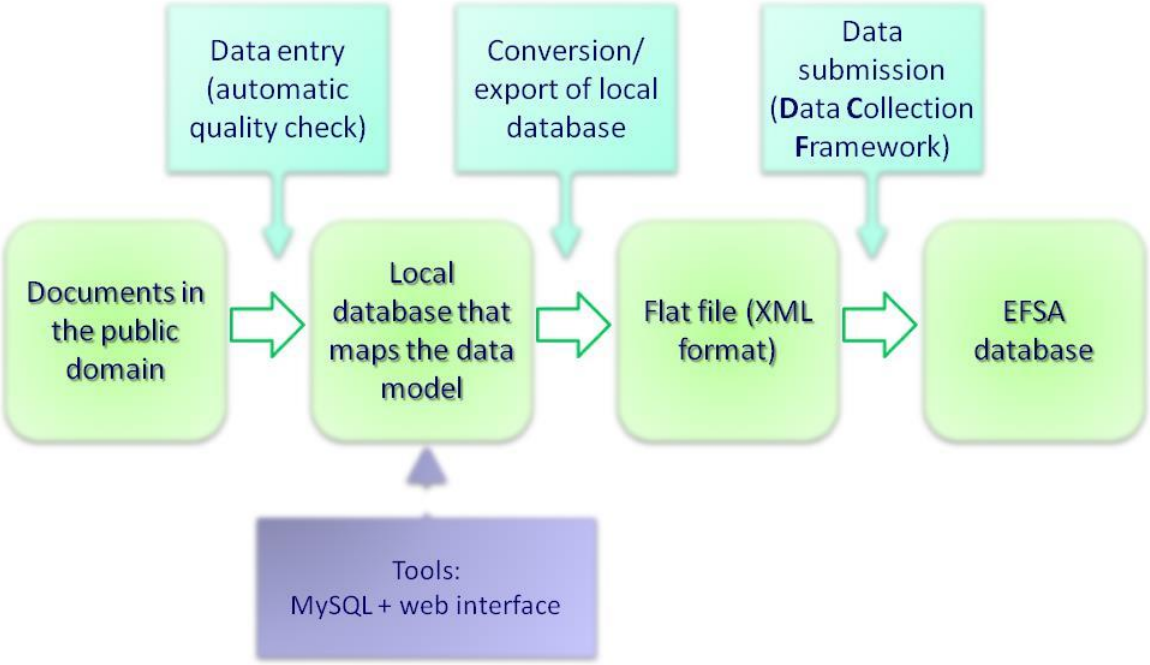

Figure 1: Overall workflow for data collection, entry and submission.

2.1.1.2 New properties

The main goal of the OpenFoodTox data model extension was to accommodate new types of information (beyond toxicity data), including physicochemical properties, PK/TK and exposure studies, available in the EFSA documents and other relevant sources.

During the first year of the project, the data model was extended to incorporate physicochemical properties and PK/TK studies. The data model was developed after thorough analysis of various types of published EFSA documents (Conclusions on Pesticides Peer Review, Scientific Opinions, Statements, Reasoned Opinions and Technical Reports) and other sources (ECHA Registered Substances Database, US FDA Pharmacological Reviews for New Drug Applications). Data availability, and the structure and format of the data reports, were considered during data model development. Applicable OECD Harmonised Templates (OHT 1 to 23-5 for physico-chemical properties and OHT 58 for PK/TK data) were also analysed and their compatibility with the data model ensured. The analysis of data sources was combined with iterative cycles of data collection. The process of data collection was established and synchronised between MN-AM and S-IN, including exchange of EFSA opinions from which hazard data were processed by S-IN into OpenFoodTox. This coordination process facilitated the retrieval of new data types from the same EFSA documents and for the same substances.

The goals of the second year of the project included:

- Further refinement and finalization of the data model after discussions and feedback from EFSA

- Update of the OpenFoodTox content by continuous collection of physico-chemical properties and PK/TK data following the previously established process of exchanging relevant EFSA documents between MN-AM and S-IN
- Extending data collection into new data sources external to EFSA
- Maintaining and ensuring the established quality of the OpenFoodTox database

The data sources included mostly published EFSA documents and US FDA Pharmacological Reviews of New Drugs Approval documents, but also a small number of evaluations of the Joint FAO/WHO Expert Committee on Food Additives (WHO/JECFA), ECHA REACH database reports, Scientific Committee Consumer Safety (SCCS) opinions, and Cosmetics Ingredient Review (CIR) assessments.

2.1.2 Quality check

2.1.2.1 Hazard data

Before each submission of the hazard data collected in the OpenFoodTox database, the following two activities are performed by S-IN: i) quality check, ii) PARAM request.

Quality check

A number of control mechanisms at the data entry level are automated to guarantee high quality of deliverables and reduce unintended errors (EFSA, 2015):

- Automatic verification of data quality at the data entry level (e.g., picklists, mandatory fields).
- Set up of rules (implemented in KNIME workflows) for automatic verification of data quality and consistency at substance, opinion, study and hazard data levels (e.g., cross-check among species, guidelines, endpoints, duration and unit for different test types).
- Manual revision of the collected data to be submitted.

Param request

For harmonization purposes, every substance/component within the OpenFoodTox 2.0 database has to be associated with a univocal PARAM code assigned by EFSA (e.g. capsorubin – RF-00009309-PAR). Each PARAM code is linked to the chemical information for that specific molecule (if available), such as CAS No., EC No., SMILES, InChI, IUPAC, molecular formula.

For chemicals without an existing PARAM code (from the EFSA PARAM catalogue), a new PARAM code should be generated. To this aim, a request for the generation of new PARAM codes is submitted to EFSA. The PARAM request consists in xls files including the list of chemicals (substances/components) without a PARAM code with the associated chemical information (i.e., name, EC no., CAS no., IUPAC, InChI, SMILES and Molecula Formula – when available), the source of that information and DOI of the most recent opinion assessig the chemical.

In the second year of the project more than 200 new chemicals (substances and /or components) were entered within the OpenFoodTox 2.0 (i.e. new chemicals assessed in EFSA documents published from May 2019). For about half of these chemicals, new PARAM codes have been requested. In some cases, a careful analysis of substance identification and composition was performed by EFSA and S-IN, in order to verify either the need for a new PARAM code generation or the assignment of an existing PARAM code.

Following the DCF submissions, a further quality check activity is performed by EFSA aimed at verifying the correctness and consistency of the hazard data extracted form the EFSA opinions (e.g., consistency among species, guidelines, exposure duration and unit for different test types). A validation report is then produced by EFSA and provided to S-IN addressing any identified discrepancy and/or clarification requests. S-IN then takes charge of the identified issues by providing clarifications to EFSA and/or revising OFT content.

2.1.2.2 New properties

The established quality of the OpenFoodTox database was maintained and ensured during new data collection (as defined above in 2.1.1.2) through a combination of automatic quality control checks at the data entry stage followed by manual revision of records deemed inconsistent. Separate QC/QA of the new content was also conducted on a specified percentage of randomly selected database records.

2.1.3 Other tools: KNIME and ACD/Name

KNIME is a modern data analytics platform that allows users to perform ETL (extract, transform, load) operations as well as sophisticated statistics and data mining. Its visual workbench combines data access, data transformation, initial investigation, predictive analytics and visualization. KNIME Desktop is open-source and available under GPL license. KNIME is extensively used to process data (e.g., cleaning, data integration) and intended to be uploaded automatically in the OpenFoodTox. Notably, KNIME includes a number of tools that can deal with molecular structures.

The chemical identification of a substance (e.g., IUPAC name, SMILES, InChI) is sometime missing or poorly reported in the EFSA documents (e.g., the structure is reported as a drawing and not codified according to standard formats such as SMILES). ACD/Name complements the identification of the chemicals whenever other publicly available resources (e.g., PubChem, ChemSpider) are not enough. ACD/Name generates chemical names according to IUPAC and CAS Index rules, converts names back to structures, and can easily handle challenging areas of nomenclature, such as biological molecules, organometallics, and polymers. The utilities that allow conversion of "name to structure" and "structure to name" prove to be vital in the data entry phase of the chemicals (including identification of stereoisomers).

2.2 Task 2/Objective 2 - design/update of new scheme for OHT results for implementing the use of EFSA overarching guidance documents

2.2.1 Case studies with OHT 201

In order to facilitate the data interoperability and straightforward exchange of information between various agencies and platforms, case studies with two tested substances (dihydroeugenol and fipronil) were set up to analyse the OECD Template #201 "Intermediate effects" and its implementation in the IUCLID 6 system. IUCLID 6 is a software tool used to record, store, submit, and exchange data on chemical substances in the format of the OHTs. The case study exercises required reporting the available information for selected "intermediate effects" with the main goals to:

i. analyze the OHT fields and their relevance;
ii. provide feedback on the OHT 201 template format in terms of the completeness of the template and controlled vocabulary pick lists;
iii. provide feedback on the data entry tool from the user perspective.

The following exercises of "intermediate effect" data reporting were conducted using IUCLID 6 software tool:

i. Test substance Dihydroeugenol using in vitro skin sensitization data provided by JRC (h-CLAT (human Cell Line Activation Test) assay);
ii. Test substance Fipronil using the data available in the scientific literature for the following endpoints: ER-alpha (estrogen receptor alpha) agonism and antagonism, TR-beta (thyroid receptor beta) agonism, antagonism, and binding/docking.

Both exercises were conducted simultaneously by participants (S-IN, MN-AM) using either stand-alone or web-based version of IUCLID 6.

3. Results

3.1 Objective 1 - Data collection for OpenFoodTox 2.0

3.1.1 Maintaining, updating hazard data in OpenFoodTox 2.0

3.1.1.1 Hazard data

The entire database includes data as extracted from the screening of more than 2140 documents (opinions, statements, conclusions) published by EFSA from 2003 to July 2020. In addition, following a request from EFSA, dietary reference values (DRV) for nutrients (minerals and vitamins) have been entered within the OFT database to harmonise OFT with the EFSA DRV finder. [3] The DRV values included AI (Adequate Intake), AR (Average Requirement), PRI (Population Reference Intake) and UL (Upper Limit) values. This activity is better described in the following paragraph. It is acknowledged that most of UL values have been extracted from SCF (EC Scientific Committee on Food) documents published before the establishment of EFSA. For this reason, the OFT database now includes also data published not only by EFSA (i.e., SCF documents) and before 2003.

Overall, the OpenFoodTox 2.0 now holds about 10300 assessments for more than 5100 chemicals (each assessment potentially including multiple (eco-)toxicity endpoints and/or hazard/risk characterisation data). About 90% of the chemical records are associated with a representative chemical structure. Figures 2-5 and Tables 1-3 summarise the contents of the OpenFoodTox. For example (see Table 2), the repository stores: 6013 assessments related to **flavourings** (2123 substances) as collected from the screening of 317 EFSA documents (opinions and statements); 1560 assessments related to **pesticides** (1266 substances) as collected from the screening of 554 EFSA documents (conclusions); 406 assessments related to **food additives** (332 substances) as collected from the screening of 223 EFSA documents (opinions and statements).

At this stage of the project, no major modifications of the standard operating procedures (SOPs) in use
for hazard data entry (Appendix A) have been performed.

The updated database can be downloaded from EFSA knowledge junction under the persistent Digital Object Identifier: https://doi.org/10.5281/zenodo.780543.

In addition, the OpenFoodTox MicroStrategy Dashboard offers a click-of-a-mouse tool to visualise the summaries of EFSA's Hazard Assessments and allows downloading of data sheets for each individual substance. It can be accessed on personal computers, smartphones and tablets through this link: https://www.efsa.europa.eu/en/microstrategy/openfoodtox.

[3] https://www.efsa.europa.eu/en/interactive-pages/drvs?lang=en

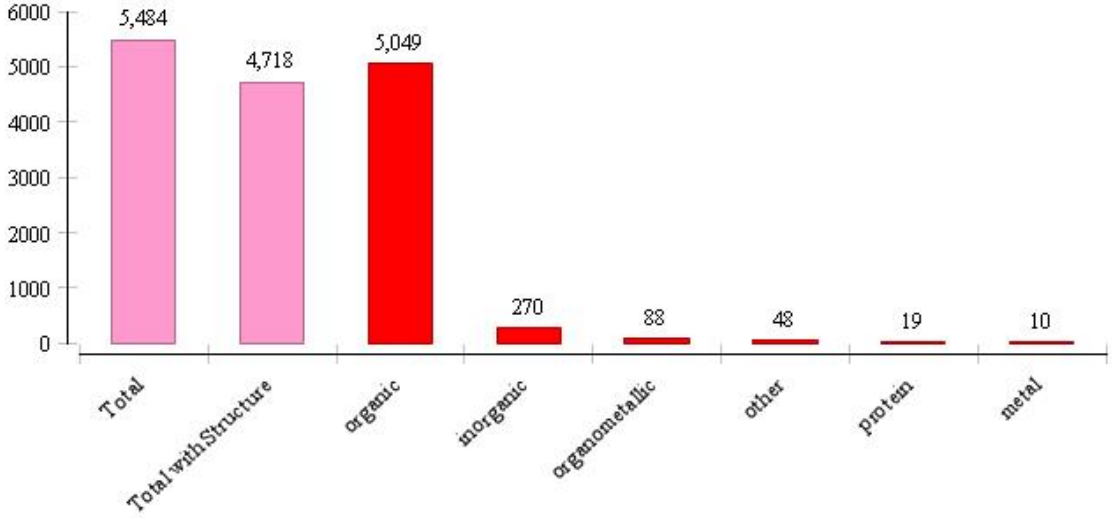

Figure 2: Number of components (and their classification) of the substances of the database. Structures in the form of SMILES (if available) are reported for components (exact or representative SMILES).

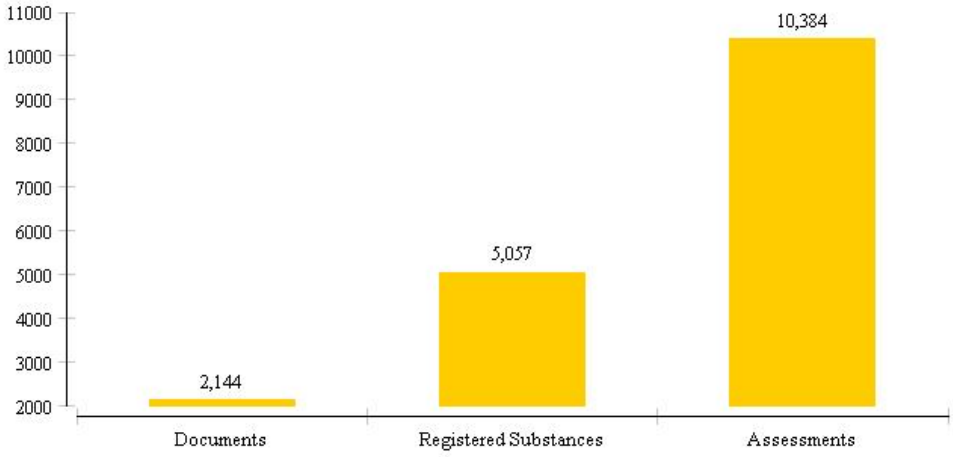

Figure 3: Number of documents (opinions/ statements/ conclusions) and substances (substances are formed by one or more components) registered in the database together with the number of assessments of a given substance as discussed in a given EFSA document.

Table 1: Classification of the EFSA's documents of the database (only documents with relevant assessments) in terms of authors (EFSA Panels). The author "EFSA" corresponds to the conclusions on the peer review of the pesticide risk assessment. *Total number of documents exceed the number of opinions as some document have two authors and they are counted twice.

Year of Publication	Author	Number of Documents
2000	SCF	5
2001	SCF	2
2002	SCF	3
2003	EFSA AFC	2
2003	SCF	4
2004	EFSA AFC	24
2004	EFSA CONTAM	9
2004	EFSA FEEDAP	17
2004	EFSA NDA	6
2004	EFSA PPR	3
2005	EFSA	15
2005	EFSA AFC	31
2005	EFSA CONTAM	12
2005	EFSA FEEDAP	10
2005	EFSA NDA	10
2005	EFSA PPR	2
2006	EFSA	36
2006	EFSA AFC	22
2006	EFSA CONTAM	4
2006	EFSA FEEDAP	7
2006	EFSA NDA	1
2006	EFSA PPR	3
2006	SCF	1
2007	EFSA	9
2007	EFSA AFC	21
2007	EFSA CONTAM	10
2007	EFSA FEEDAP	11
2007	EFSA NDA	2
2007	EFSA PPR	2
2008	EFSA	37
2008	EFSA AFC	62
2008	EFSA ANS	11
2008	EFSA CEF	4
2008	EFSA CONTAM	22
2008	EFSA FEEDAP	11
2008	EFSA GMO	2
2008	EFSA NDA	4
2008	EFSA PPR	1
2009	EFSA	54
2009	EFSA AFC	20
2009	EFSA ANS	44
2009	EFSA CEF	42
2009	EFSA CONTAM	13
2009	EFSA FEEDAP	13
2009	EFSA NDA	3
2009	EFSA PPR	2
2010	EFSA	66
2010	EFSA AFC	1
2010	EFSA ANS	33
2010	EFSA CEF	37
2010	EFSA CONTAM	11

2010	EFSA FEEDAP	12
2010	EFSA GMO	1
2010	EFSA NDA	8
2011	EFSA	33
2011	EFSA ANS	13
2011	EFSA CEF	67
2011	EFSA CONTAM	12
2011	EFSA FEEDAP	32
2011	EFSA NDA	5
2012	EFSA	50
2012	EFSA ANS	13
2012	EFSA CEF	40
2012	EFSA CONTAM	10
2012	EFSA FEEDAP	54
2012	EFSA NDA	8
2012	EFSA PPR	1
2013	EFSA	45
2013	EFSA ANS	19
2013	EFSA CEF	35
2013	EFSA CONTAM	3
2013	EFSA FEEDAP	52
2013	EFSA NDA	11
2013	EFSA PPR	1
2014	EFSA	38
2014	EFSA ANS	9
2014	EFSA CEF	34
2014	EFSA CONTAM	4
2014	EFSA FEEDAP	39
2014	EFSA NDA	11
2015	EFSA	32
2015	EFSA ANS	26
2015	EFSA CEF	30
2015	EFSA CONTAM	6
2015	EFSA FEEDAP	36
2015	EFSA NDA	16
2016	EFSA	34
2016	EFSA ANS	22
2016	EFSA CEF	22
2016	EFSA CONTAM	5
2016	EFSA FEEDAP	50
2016	EFSA NDA	11
2017	EFSA	23
2017	EFSA ANS	21
2017	EFSA CEF	19
2017	EFSA CONTAM	7
2017	EFSA FEEDAP	24
2017	EFSA NDA	15
2018	EFSA	41
2018	EFSA ANS	25
2018	EFSA CEF	11
2018	EFSA CEP	3
2018	EFSA CONTAM	9

2018	EFSA FAF	2
2018	EFSA FEEDAP	32
2018	EFSA NDA	9
2019	EFSA	9
2019	EFSA CEF	1
2019	EFSA CEP	6
2019	EFSA CONTAM	3
2019	EFSA FAF	19
2019	EFSA FEEDAP	54
2019	EFSA NDA	8
2019	EFSA PPR	4
2020	EFSA CEP	4
2020	EFSA CONTAM	3
2020	EFSA FAF	13
2020	EFSA FEEDAP	35
2020	EFSA NDA	10
2020	EFSA PPR	13

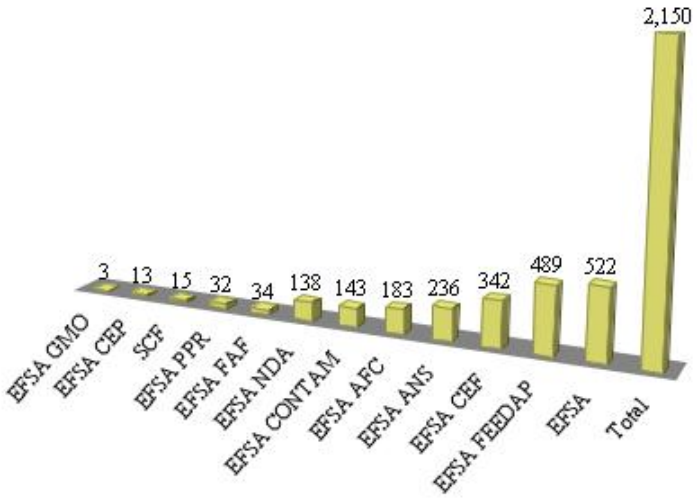

Figure 4: Classification of the EFSA's documents of the database in terms of their authors. The author "EFSA" corresponds to the conclusions on the peer review of the pesticide risk assessment. Total number of documents exceedes the number of opinions as some documents have two authors and they are counted twice.

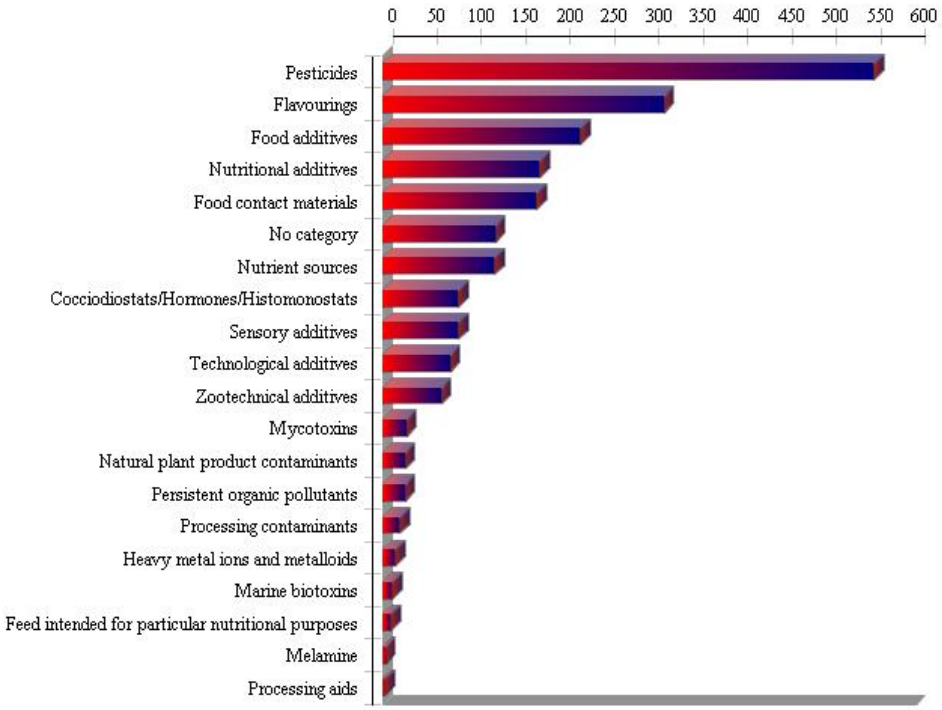

Figure 5: Classification of the EFSA's documents of the database in terms of subareas.

Table 2: Number of EFSA documents (opinions, statements, conclusions), number of assessments (assessment of a substance in a given opinion), and number of substances for the different subareas. Substances may be discussed and assessed in more than one document; a document of a given subarea may discuss and deal with more than one substance.

Category	Number of documents	Number of assessments	Number of substances
Cocciodiostats/Hormones/Histomonost	85	122	54
Feed intended for particular nutritional purposes	9	18	17
Flavourings	317	6013	2123
Food additives	223	406	332
Food contact materials	174	397	341
Heavy metal ions and metalloids	14	25	16
Marine biotoxins	11	76	76
Melamine	3	8	5
Mycotoxins	27	77	63
Natural plant product contaminants	25	50	40
No category	127	191	163
Nutrient sources	126	292	195
Nutritional additives	177	266	182
Persistent organic pollutants	25	106	91
Pesticides	554	1560	1266
Processing aids	2	7	7
Processing contaminants	19	107	100
Sensory additives	85	479	467
Technological additives	76	96	61
Zootechnical additives	66	88	60

Table 3: Summary of the hazard/risk characterisation records with focus on populations.

Population	Types of hazard/risk assessment	Total number of records
Aquatic Invertebrates	margin of safety	2
Aquatic Vertebrates	critical study not identified |maximum safe intake/maximum safe concentration in feed |margin of safety |maximum tolerated level/dose |TTC Cramer Class II |TTC Cramer Class I |TTC Cramer Class III	895
Aquatic compartment	PNEC |margin of safety	100
Human	critical study not identified |TTC genotoxicity |margin of safety |UL |ADI |TDI |TDI, provisional (PTDI) |TDI, provisional maximum (PMTDI) |TWI |group TDI |MOE |RfD |TWI, provisional (PTWI) |group ADI |maximum safe intake/maximum safe concentration in feed |MoBB |TTC Cramer Class III |ARfD |UL, provisional (PUL) |OSL |ADI, provisional |group TWI |TTC Cramer Class II |TTC Cramer Class I |AOEL |AOEL, provisional |AOEC, provisional |MTDI |group ARfD |AAOEL |safe maximum intake level (in food) |AI (Adequate Intake) |PRI (Population reference intake) |AR (Average requirement) |AOEC |AAOEC	13407
Sediment compartment	PNEC |margin of safety	100
Soil compartment	PNEC |margin of safety	78
Terrestrial Plants	PNEC	6
Terrestrial Vertebrates	critical study not identified |maximum safe intake/maximum safe concentration in feed |margin of safety |maximum tolerated level/dose |TTC Cramer Class II |TTC Cramer Class I |TTC Cramer Class III |UL |MOE	4111

3.1.1.2 DRV data

As anticipated in the previous paragraph, dietary reference values (DRV) for a subset of nutrients (minerals and vitamins) have been entered within the OFT database. Nearly 500 DRV values, which are publicly available through the EFSA DRV finder interactive tool, have been provided by EFSA as xlsx file together with the links to the original documents in which these values have been established (ca. 50 documents). The DRV values included AI (Adequate Intake), AR (Average Requirement), PRI (Population Reference Intake) and UL (Upper Limit) values.

The DRV data provided by EFSA (xlsx file) have been analysed in detail i) to check for consistency with the hazard data already included in OFT, ii) to check for consistency with the SCF/EFSA original documents, and iii) to adapt the information extracted from the DRV finder with the data model of the OFT database. As an example, new target populations have been added in the EFSA catalogue (and in OFT) to allow the match between the two databases. In addition, much attention was paid to assign the proper PARAM code to the substances (and related components) for which the DRV values have been defined, i.e. minerals and vitamins, to assure that the DRV values included in the OFT database were referred to the substances assessed in the SCF/EFSA documents.

This activity was concluded by the beginning of February 2020 and all DRV data have been included in the DCF submission performed on 20th February 2020.

3.1.2 Modifications of the OpenFoodTox data model using OECD Harmonised Templates

3.1.2.1 Data modelling process

During the first year of the framework contract, a cyclic process of data harvesting and modelling produced formalized data models with respect to physicochemical data as well as for PK/PK and ADE data. This process was continued in year two, and corresponding data models were further refined. Figure 6 depicts a high-level summary of the data modelling approach employed by MN-AM.

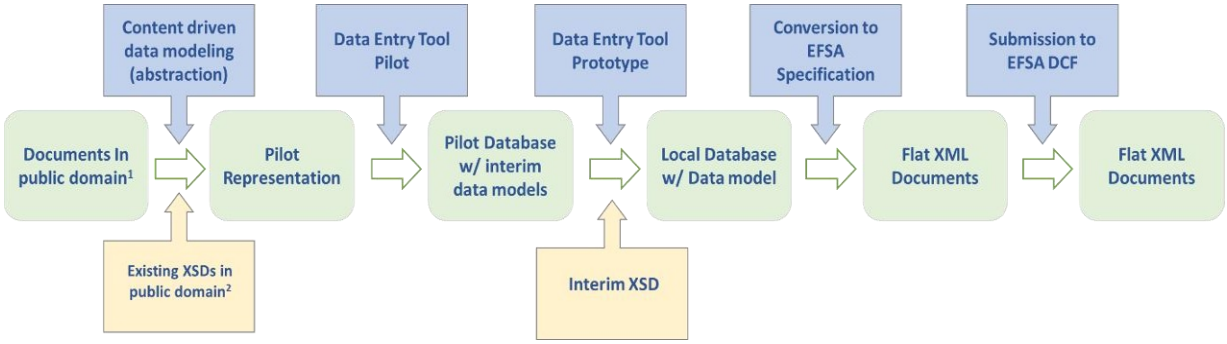

Figure 6: Summary of process for addition of new data. 1 Analysed documents in public domain included: EFSA documents, dossiers of registered substances from ECHA database, US FDA pharmacological reviews for new drug application documents; 2 Existing data models for physicochemical properties (EFSA), PK/TK (ECHA, EFSA, US FDA), and ADME studies (ECHA, EFSA, US FDA).

Data models for both physicochemical and PK/TK properties are based on a reusable, custom XML format which records property values and conditions and possess the versatility to capture various data types. Of note, the format was designed to comply with numerical and textual values of various unnormalized specifications. A recently updated unified definition of "property value" (see Figure 7) incorporated into the data model can be applied to a variety of specified values such as:

- Optional qualifier (>, <, etc.)
- Numeric value with optional fields:
  - Minimum
  - Maximum
  - Standard deviation
  - Mean value (new)

    - Geometric mean value (new)
    - Geometric standard deviation (new)
    - Standard error (new)
- Text value
- Optional unit
- Optional remarks

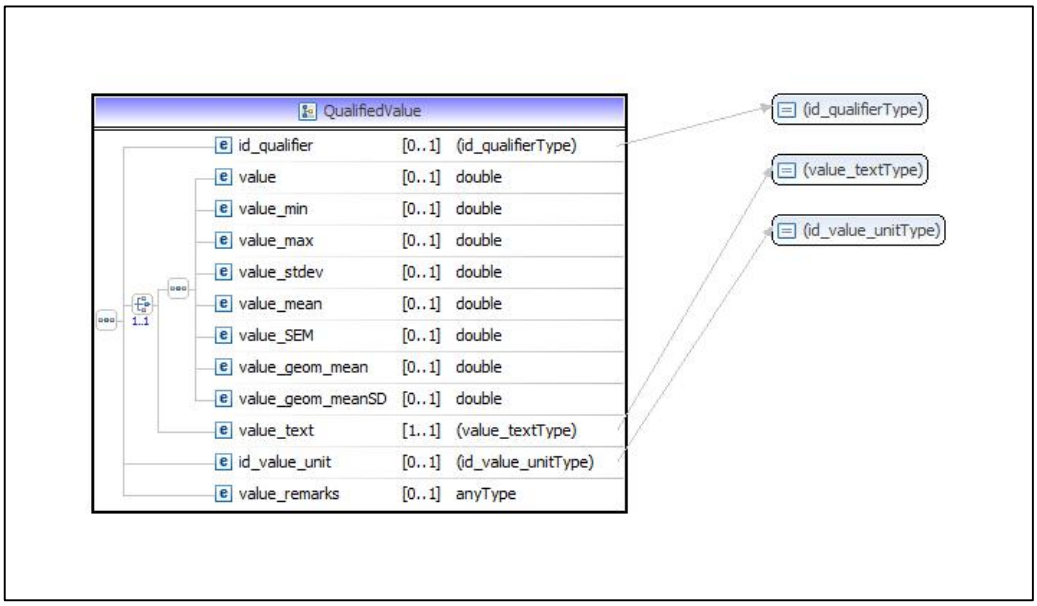

Figure 7: Updated custom type for qualified values

3.1.2.2 Data model: Physicochemical properties

Harvested physico-chemical property data exists in two forms, namely, experimental and calculated properties. To account for this fact, the corresponding data model table was split into two separate entities, experimental physico-chemical properties and calculated physicochemical properties, as shown in Figures 8 and 9, respectively. Furthermore, both tables were also supplemented with several additional metadata fields.

Regarding experimental physicochemical properties, for each recorded measurement appropriate experimental conditions were captured including temperature, pressure, pH, wavelength and concentrations. The information on the compliance (and/or deviations) of the experiments with (from) relevant guidelines and/or GLP was also captured when available. The measurement method as well as Limit of Detection (LoD) and Quantification (LoQ) are specified. Each tested substance was characterized in terms of its composition (with PARAM codes for substance and components), tested form, and purity (%).

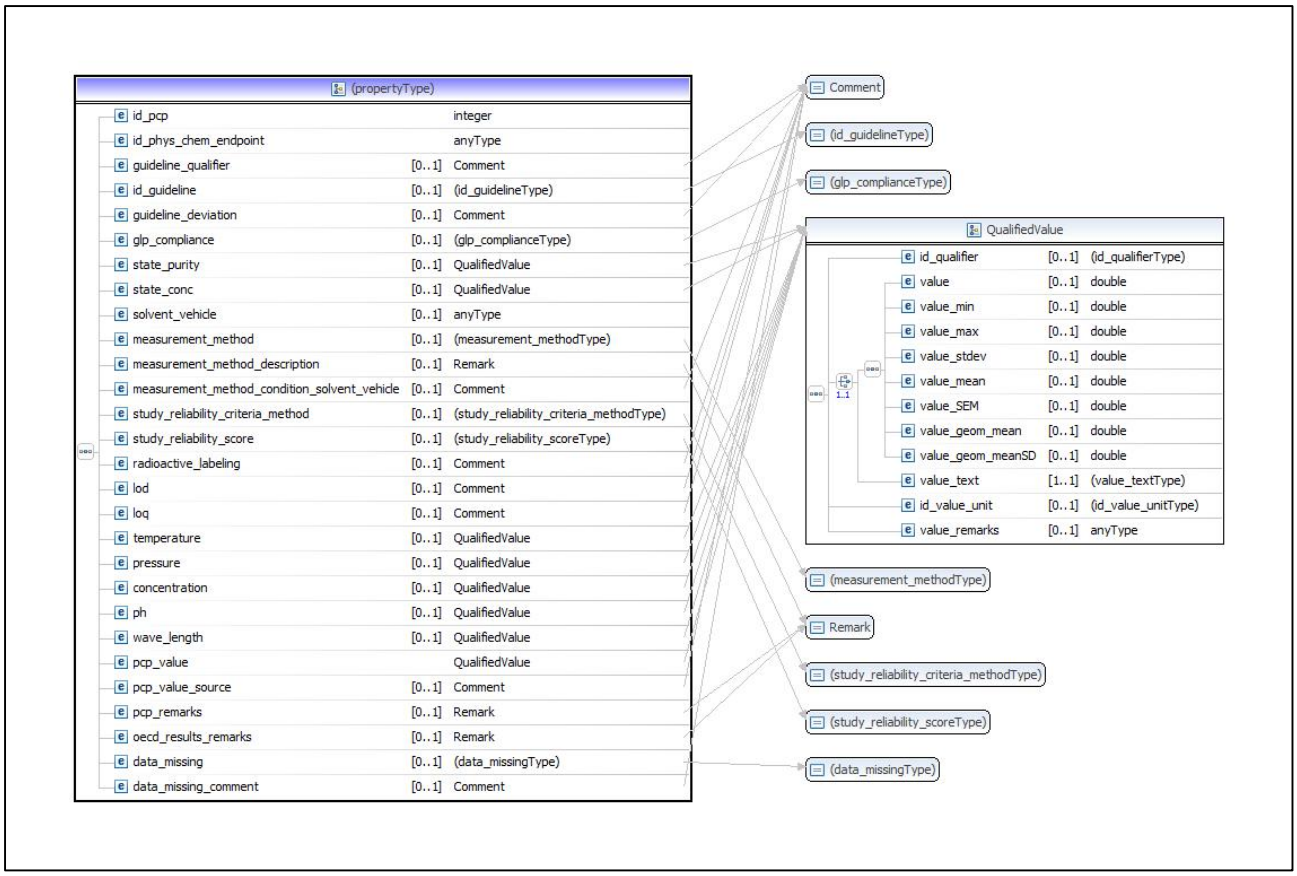

Figure 8: The XSD model for experimental physico-chemical properties.

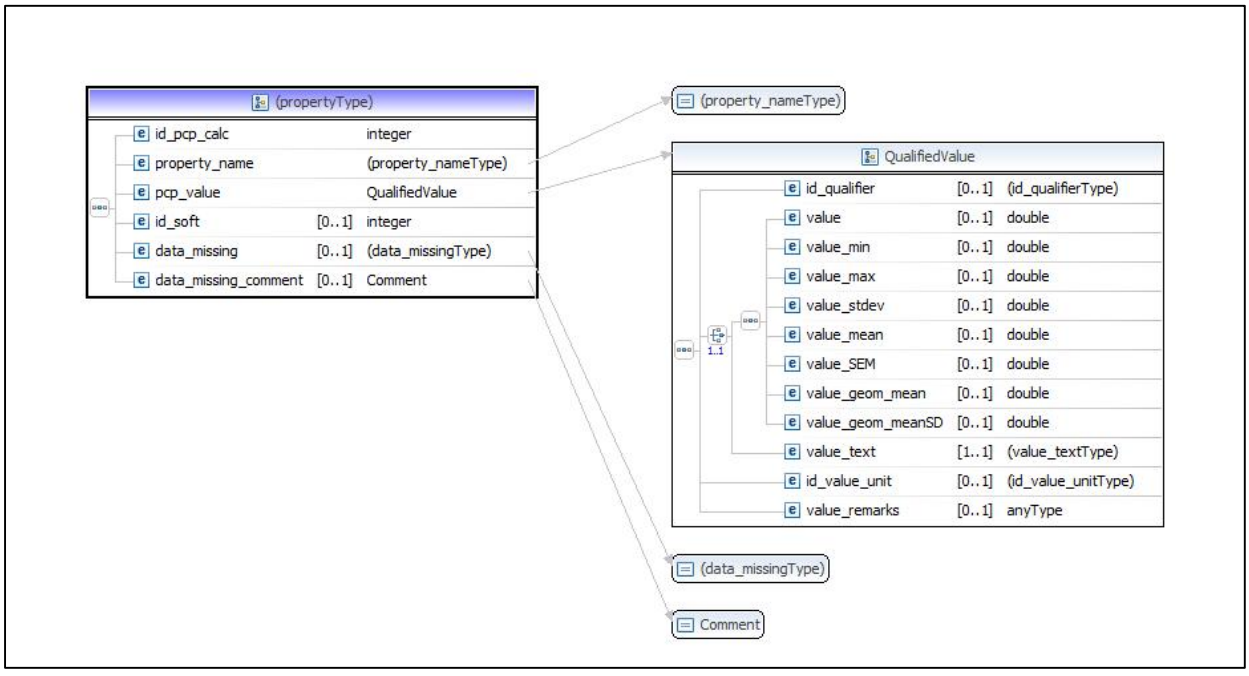

Figure 9: The XSD model for calculated physico-chemical properties.

Corresponding XSD schema files make use of the custom XML format to record various types of property values.

Regarding calculated physico-chemical properties in particular, a separate table for recording software used for calculations was also introduced as shown in Figure 10. This table includes information such as software name, developer and version, as well as text field for any relevant coments related to the calculation conditions (e.g., water solubility at 25 degC).

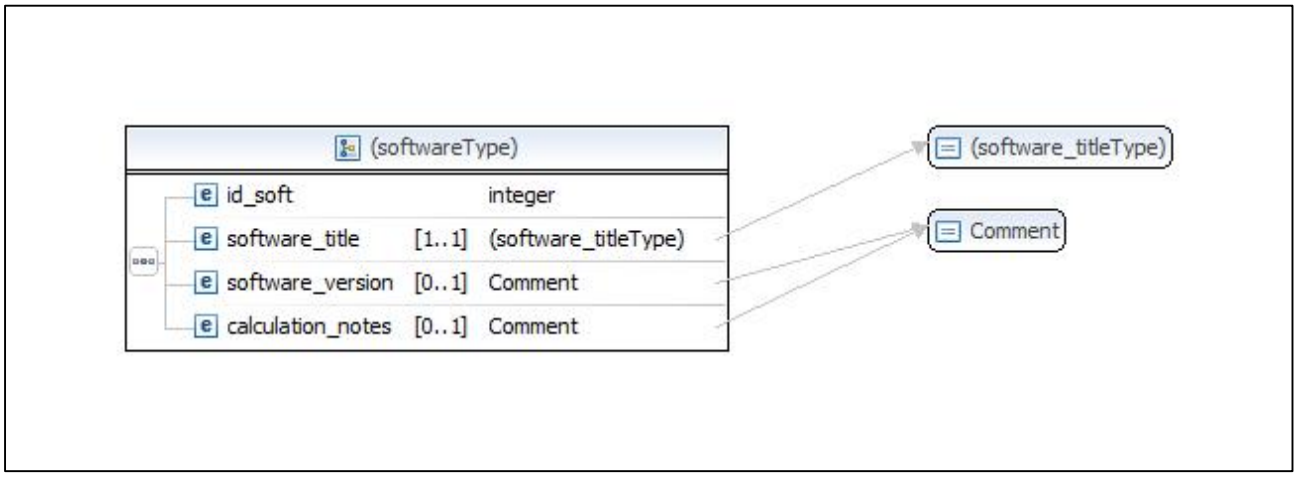

Figure 10: The XSD model for software used for calculation of physico-chemical properties.

3.1.2.3 The XSD data models for PK/TK and ACE data

Similar to the physico-chemical data models, the PK/TK data model is also based on the custom XML format designed to record various types of property values and conditions, as well as annotations for reported values (shown on Figure 11).

For each PK/TK assay, the study design information was recorded including test animal information (species, sex, number of animals in test and control groups, animals age and/or weight), route of administration, study duration, doses, and dosing regimen. The captured information on analytical methods include method type, Limit of Detection (LoD) and Limit of Quantification (LoQ). Each test substance was characterized with respect to its composition (with PARAM codes for substances and components), purity, tested form, and radiolabeling. The study reliability as well as GLP-/GCP- (Good Laboratory/Clinical Practice) and guideline compliance/deviations were also recorded.

As a result of data collection conducted during the second year of the project and the expansion of the data sources beyond EFSA, and following the discussions with EFSA, the PK/TK and ADE data model was supplemented with additional fields to capture additional information on experimental conditions, study design details, analytical methods, test substances characterization, and study reliability, as available in Figure 11.

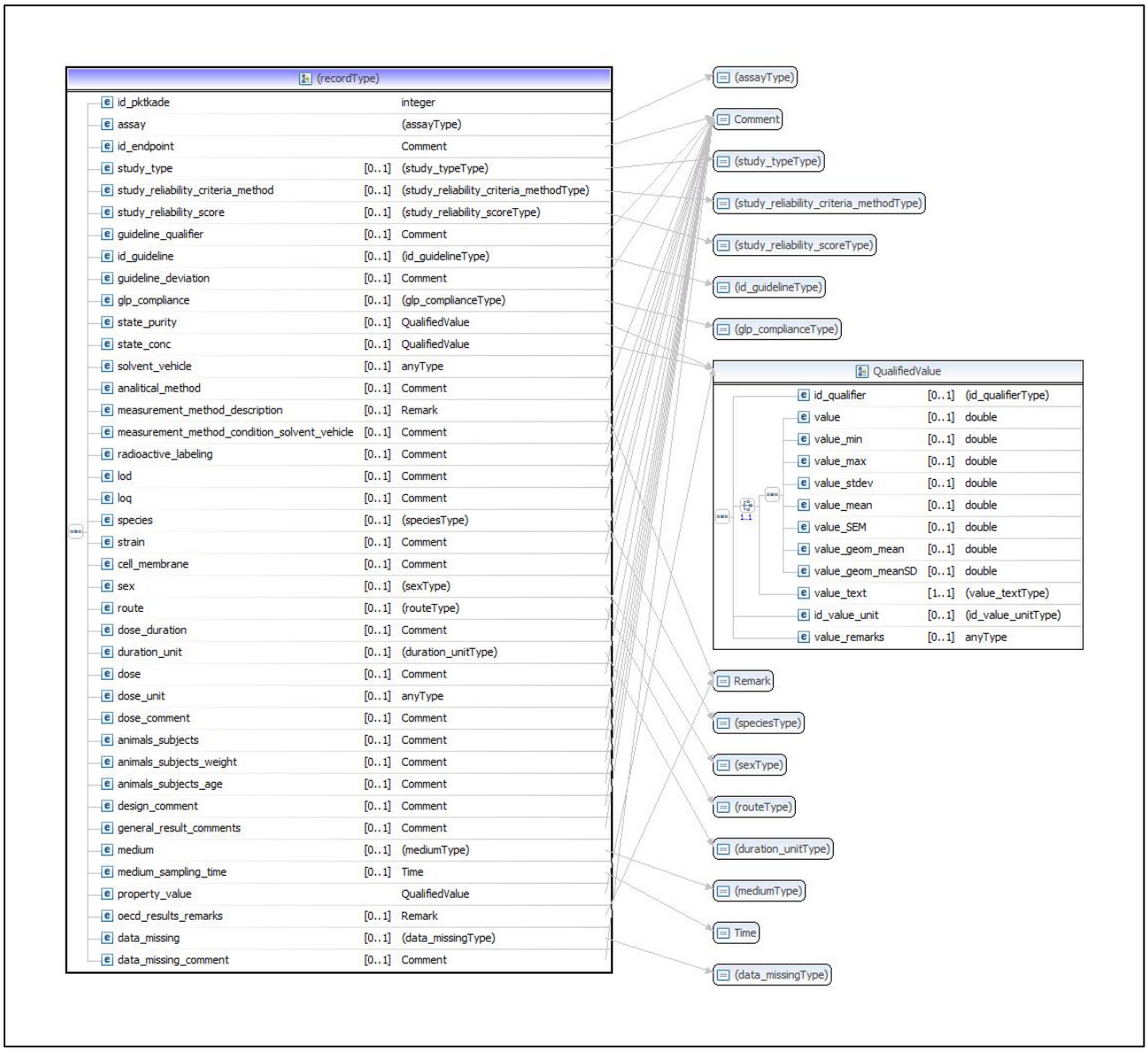

Figure 11: The XSD model for PK/TK data.

Data models for both physico-chemical properties and PK/TK data document the XSD extensions and support the development of new catalogues for newly defined pick-lists.

3.1.3 Data collection of new substance properties

The process of data collection established during the first year of the project was continued and extended in the second year. Recent EFSA opinions processed by S-IN into OpenFoodTox were considered for the extraction of new physico-chemical and PK/TK property data. Moreover, as explained in greater detail in section 3.1.3.2, data sources outside of EFSA were considered. The quality of the OpenFoodTox database was ensured during the collection of new data through a combination of automatic quality control checks at the data entry stage followed by manual revision of records identified as inconsistent. A separate QC/QA operation of the new content was also conducted on a specified percentage of randomly selected database records.

3.1.3.1 Physico-chemical properties

In the second year of the project, 65 EFSA documents covering 201 substances (240 components) were searched for physico-chemical property data. The relevant information was identified in 45 opinions, including 183 tested substances, mostly "Food Additives and Flavourings", and "Additives and Products or Substances used in Animal Feed" (Table 4). No data on experimental physico-chemical properties was found in 20 of the analysed opinions.

Table 4: Overview of substance types with experimental physico-chemical property data collected in Year 1 and Year 2 of the project including counts of substances with data and corresponding EFSA documents.

EFSA PANEL	Substance types	# Substances with data | # EFSA Documents	
		Year 1	Year 2
EFSA CEP*	Food Contact Materials, Enzymes and Processing Aids	155 | 10	3 | 3
EFSA CONTAM	Contaminants in the Food Chain	15 | 5	2 | 1
EFSA FAF**	Food Additives and Flavourings	82 | 20	140 | 12
EFSA FEEDAP	Additives and Products or Substances used in Animal Feed	36 | 32	30 | 21
EFSA NDA	Novel Foods and Food Allergens	3 | 2	2 | 2
EFSA/EFSA PPR	Plant Protection Products and their Residues	48 | 24	7 | 6
TOTAL		**339 | 93**	**183 | 45**

* including previous CEF ** including previous ANS

Experimental data was recorded for 23 property types (see Figure 12). Solubility in water and organic solvents, as well as melting and boiling points, were the most frequently reported properties. Appearance data included information on colour, odour and the physical state of tested substances. Calculated properties (e.g. Henry's law constant and partition coefficients) and details on the calculation method and/or software tool used were also recorded, when available. Idenified data gaps were also recorded in the database.

The number of database records collected during Year 1 and Year 2 of the project and corresponding substance counts are presented in Table 5.

Table 5: Summary statistics of collected physico-chemical property and appearance data.

Data type	# Database records with data | # Substances with data	
	Year 1	Year 2
Appearance	487 | 309	332 | 188
Physicochemical properties - calculated	32 | 26	3 | 2

Physicochemical properties - experimental	2413 | 339	1245 | 183

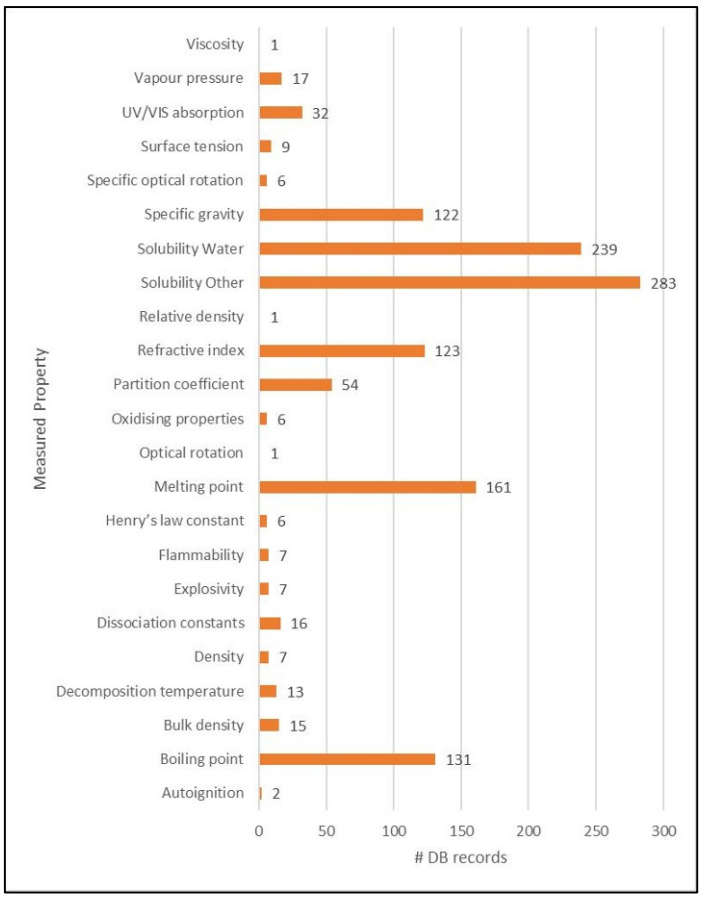

Figure 12: The number of database records available for experimentally-derived physicochemical properties.

3.1.3.2 Pharmacokinetic/Toxicokinetic data

During the first year of the project, several data sources (including EFSA documents and Draft Assessment Reports (DARs), US FDA Pharmacological Reviews for New Drug Applications (NDAs), the ECHA Registered Substances Database, and EU SCCS opinions) were analysed for the purpose of incorporating PK/TK data into the model design. The actual data was collected from 58 EFSA documents. Due to the complexity of the data and significant differences in the reporting methodology of various opinions, the extraction process was labour intensive. At the same time, it was observed that data available via EFSA documents is provided mainly in descriptive form, and only a relatively small number of quantitative data could be collected. This notion was especially true for substances other than pesticides. Furthermore, dose-level information was rarely provided in EFSA opinions. Therefore, the data collection was extended to sources other than EFSA in the second year of the project in order to extract quantitative toxicokinetics parameters, preferably with dose-level information, whenever possible.

To this end, the data for 126 "EFSA substances", expressed as substance-component pairs as defined by the chemistry data model of OFT databse, were obtained from 50 EFSA opinions. All substances were reviewed by EFSA and consists mostly of "Food Additives and Flavourings" and "Additives and Products or Substances used in Animal Feed", as presented in Table 6. Data gaps for particular endpoints were also identified.

Table 6: Overview of substance types with PK/TK data collected from EFSA documents in Year 1 and Year 2 of the project including the number of substances with available data and the number of corresponding EFSA documents.

EFSA PANEL	Substance types	#Substances with data*** | # EFSA Documents	
		Year 1	Year 2
EFSA CEP*	Food Contact Materials, Enzymes and Processing Aids	80 | 4	1 | 1
EFSA CONTAM	Contaminants in the Food Chain	14 | 2	9 | 1
EFSA FAF**	Food Additives and Flavourings	30 | 11	82 | 22
EFSA FEEDAP	Additives and Products or Substances used in Animal Feed	30 | 13	23 | 20
EFSA NDA	Novel Foods and Food Allergens	5 | 4	6 | 2
EFSA/EFSA PPR	Plant Protection Products and their Residues	30 | 24	5 | 4
TOTAL		**189 | 58**	**126 | 50**

*including previous CEF and AFC ** including previous ANS *** expressed as substance-component pairs as defined by the chemistry data model of OpenFoodTox database

The collected toxicokinetics data for "non-EFSA substances" (i.e. substances not reviewed by EFSA) covered 60 drugs (from US FDA NDA documents) and 6 cosmetic ingredients (using WHO/JECFA, ECHA, SCCS, and CIR documents). Of particular note, data sourced from US FDA NDAs included well-documented studies with exhaustive design details and numeric data for each dose-level tested.

Table 7 presents an overview of the database content including the number of study records available for considered assays (absorption/bioavailability, excretion, distribution, bioaccumulation, and toxicokinetics) and the corresponding number of covered tested substances.

Table 7: Summary statistics of pharmacokinetic/toxicokinetic data collected in Year 1 and Year 2 of the project.

PK/TK ASSAY	# Database records with data | # Substances* with data	
	Year 1	Year 2
Absorption/Bioavailability	201 | 154	159 | 76
Excretion	243 | 143	268 | 76
Distribution	410 | 134	337 | 46
Bioaccumulation	37 | 34	27 | 18
Toxicokinetics	115 | 46	854 |151

*Expressed as substance-component pairs as defined by the chemistry data model of OpenFoodTox database

Regarding the data content collected in Year 2 of the project, the most populated species across all recorded assays (for all substances) were rodents, followed by primates, poultry, and pigs. The dominating route of exposure was oral (Figure 13).

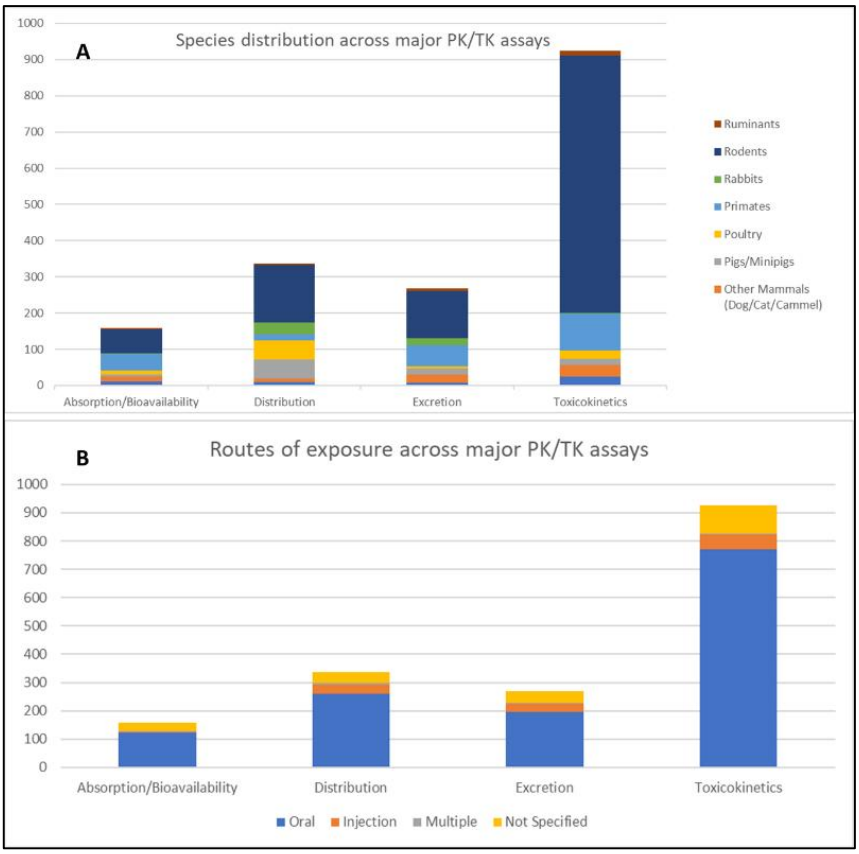

Figure 13: Species (A) and routes of exposure (B) covered by particular PK/TK assays.

As mentioned, special emphasis was placed on collecting quantitative data for key toxicokinetics parameters, such as AUC, Cmax, T-max, and T-half. The current status of available quantititive data in the databse is presented in Table 8. As demonstrated, the largest number of data was available for drugs from US FDA NDAs.

Table 8: Availability of quantitative data for analysed PK/TK assays including the number of data records and unique test substances.

PK/TK Assay	Parameter	# Numeric Data Records | # Substances	
		Year 1	Year 2
Absorption / Bioavailability	Oral absorption (% and/or amount)	89 | 55	37 | 21
	Systemic bioavailability (% and/or amount)	33 | 19	22 | 14
Distribution	Distribution Amount (% and/or amount)	147 | 11	137 | 18
Excretion	Total; Urinary; Biliary; Faecal; Exhaled air; Amount in carcass; Amount in cage wash (% and/or amount)	121 | 44	161 | 50
Toxicokinetics – "EFSA substances"	Cmax	55 | 21	39 |15
	T-max	68 | 28	48 | 15
	AUC	39 | 11	38 | 17

	T-half	65 | 26	63 | 15
Toxicokinetics – "non-EFSA substances" (drugs) – dose-level data	Cmax	N/A	221 | 29
	T-max	N/A	120 | 17
	AUC	N/A	240 | 27
	T-half	N/A	49 | 10

Distribution data was recorded for 54 sites (including organ- and tissue- levels). The most frequently reported distribution organ was liver, followed by kidney, muscle, fat and brain. In a similar manner to Year 1, the potential for bioaccumulation was recorded in a descriptive manner, and as such there were no numeric values provided in the harvested EFSA documents for this assay.

3.2 Objective 2 - Design/update of OHT 201 using case studies

The following "intermediate effect" data, summarised in Table 9, were entered using OHT 201 template available via IUCLID 6 software tool:

i. in vitro skin sensitization data (human Cell Line Activation Test) for dihydroeugenol;
ii. ER-alpha agonism and antagonism, TR-beta agonism and antagonism, and binding/docking literature data for fipronil.

The export of the data from IUCLID 6 stand-alone version using the classic UI is provided in Appendix B, C and D. The combined feedback from all participants can be summarized as follows:

- The IUCLID 6 fields and controlled vocabulary picklists are fully compliant with OHT 201 and no missing items were found.
- For the new users (who do not have experience with entering the data using IUCLID 6) navigation through the tool (e.g., finding the proper harmonized template) may be challenging and require consulting the documentation and spending a lot of time to get familiar with the software. This would however not be an issue for more experienced user.
- Editing added records (e.g., test material or references) and navigating the tool after edits (for example, returning to the OHT 201 after editing the information about the test material) is not straightforward.
- Text fields are very general (e.g.: details on test system) and equipping the template with more specific text fields could facilitate data submission.
- The table of contents is not exported correctly from the IUCLID 6 tool.

Table 9: Summary of the intermediate effects used for the OHT201 case study.

Substance	Intermediate effect	Imputed value	Overall conclusion	Ref.
Dihydroeugenol	hCLAT: Human monocytic leukemia cell line (THP-1 cells)	EC150 for CD86: RFI[1] = 195 (at dose 162 micro-g/mL) (the only positive run)	Skin sensitizer	Ashikaga T et al. Toxicol In Vitro 20(5): 767-73
Fipronil	ER-alpha agonism: CHO-K1 (Hamster Ovary) cells	REC20 (M)[2] = No effect	Fipronil did not act as ER alpha agonist	Lu M. Chemosphere 120 (2015) 246–251
	ER-alpha antagonism: CHO-K1 (Hamster Ovary) cells	REC20 (M)[2] = 0.00000064 RLAb (%)[3] = 77	Fipronil showed agonistic activity - inhibited the estrogenic activity via ER alpha.	Lu M. Chemosphere 120 (2015) 246–251

			Also, the luciferase activity was significantly inhibited	
	TR-beta agonism: CHO-K1 (Hamster Ovary) cells	REC20 (M)[4] = No effect	Fipronil did not act as TR beta agonist	Lu M. Chemosphere 120 (2015) 246–251
	TR-beta antagonism: CHO-K1 (Hamster Ovary) cells	REC20 (M)[4] = No effect	Fipronil did not act as TR beta antagonist	Lu M. Chemosphere 120 (2015) 246–251
	Binding value: Prediction of binding modes of fipronil to TR-beta using molecular docking performed with MOE 2009 and Molegro Virtual Docker (MVD version 4.2) software tools	MolDock Score[5] = -117.007	Fipronil was well fitted at the binding site of the receptor and embedded in a binding pocket composed of Asn331, Ala 279, Asn 233, Arg 316 Arg 317 and Leu 330	Lu M. Chemosphere 120 (2015) 246–251

[1] RFI - Relative Fluorescence Intensity. RFI values > 200 for CD54 and > 150 for CD86 are considered as a positive response of these cell surface makers that lead to a classification as skin sensitizer, [2]REC20 - the concentration of the test chemicals showing 20% of the agonistic activity 2x10[-8] M of 17beta-estradiol (positive control) via ER alpha. [3] RLAb (%) - Relative Luciferase Activity - percentage response of maximum inhibition of the test chemical with 100% activity defined as the activity achieved to 1x10[-9] M of 17beta-estradiol (positive control). [4] REC20 - the concentration of the test chemicals showing 20% of the agonistic activity of 1x10[-7] M of a positive control (T3) via TR beta. [5] MolDock Score - Molecular docking score for TR-beta between fipronil.

3.3 Objective 3 - Designing an in silico integrative tool allowing description and prediction of hazard properties of chemicals for OpenFoodTox 2.0

Ten different QSARs models have been developed for new toxicological endpoints in species of ecological relevance such as earthworm, Daphnia magna, rainbow trout, and honey bees. In addition, human toxicology data were collected to build a model for inhalation toxicity. Finally, a tool to classify substances according to mode of actions has been obtained.

Data on hazard properties have been collected from an updated version of EFSA's OpenFoodTox (January 2020). In particular, the following QSAR models have been developed according to different approaches:

a. One regression-based model (n=233) on pesticides acute toxicity (pLC50) in rainbow trout (O. mykiss) (Toropov et al., 2020);
b. One regression-based model (n = 58) on pesticides acute toxicity (LC50) in earthworm (E. fetida) (Ghosh et al., 2020);
c. One classification-based models (n = 163) on pesticides acute toxicity (LC50) in earthworm (E. fetida) (Roy et al., 2020)
d. Two regression-based models (n = 123 and n = 97, respectively) on binary mixtures potency (LD50-mix) (Carnesecchi et al., 2020);
e. Integrative QSAR model to predict acute contact toxicity (LD50) and to profile the Mode of Action (MoA) of pesticides active substances in honey bees (*A. mellifera*) (Carnesecchi et al., 2020a);
f. Regression-based models on biocides acute toxicity (EC50) in freshwater crustacean *Daphnia magna* (Marzo et al., 2020);
g. Regression-based models for acute toxicity of pesticides towards *Daphnia magna* (Cappelli et al., 2020)
h. Regression- and classification-based models on pesticides acute toxicity (LC50) in bobwhite quail (C. virginianus) (Carnesecchi et al. in preparation).
i. Regression- based model for inhalation toxicity as No Observed Adverse Effect Concentration (NOAEC) (Toropov et al., submitted)
j. Regression-based model to predict water solubility of pesticides (Toropova et al., 2020)

a. Regression based model – rainbow trout

In this study, a quantitative structure-activity relationship (QSAR) approach was established for the prediction of acute toxicity toward rainbow trout of various pesticides. The endpoint is expressed as the negative logarithm of lethal concentration (mM/L).

The index of ideality of correlation (IIC) is the main component of this approach. IIC identifies the best models, measuring the deviations from the training set and the invisible training set, to avoid overfitting. The advantage of the IIC is its sensitivity to two statistical characteristics which are usually analyzed separately, i.e. to the correlation coefficient and to mean absolute error (MAE).

The specific model applied to pesticides, which are toxic and often have complex structures, shows good performance particularly on the external validation set of compounds, suggesting it is applicable to the prediction of new substances. Another advantage is that the model is quite simple and only needs the SMILES structure as input.

Table 10: Statistical quality of the rainbow trout model.

$n_{training}$	$r^2_{training}$	$n_{validation}$	$r^2_{validation}$
233	0.67	78	0.86

b. Regression based model - earthworm

This model has been described in the previous report. In 2020 we published the paper regarding this model, containing the full details of the model. Figure 14 shows the results of the predicted versus experimental values.

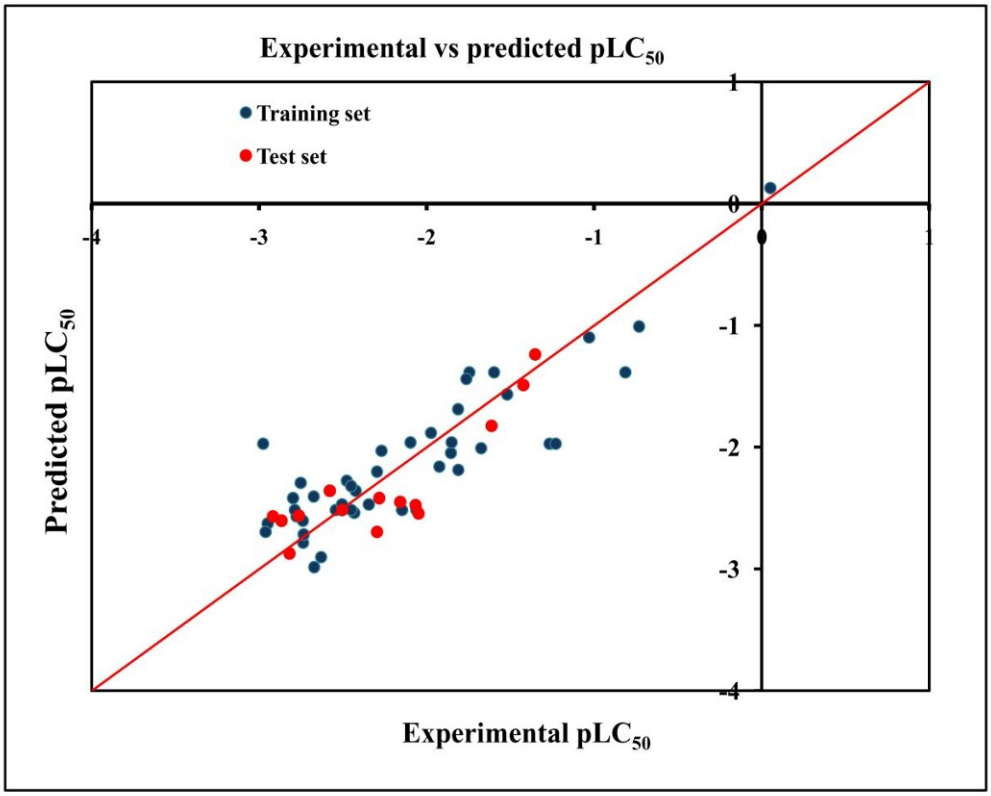

Figure 14. The scatter plot of the observed and the predicted pLC50 values obtained by the developed PLS model.

c. Classification based model - earthworm

This model has been described in the previous report. In 2020 we published the paper regarding this model, containing the full details of the model. The best discriminant model obtained with 8 descriptors showed appreciable Wilks' λ value of 0.490, F (Fischer's statistics) value of 14.03, χ2 value of 79.098, canonical regression coefficient (R) value of 0.714 and ρ value of 14.63. The sensitivity, specificity, accuracy, precision and F-measure values of the training set are 90.00, 80.52, 83.76, 70.59 and 79.12 respectively whereas for the test set, these are 58.82, 79.31, 71.74, 62.50 and 60.61 respectively.

The insights obtained from the LDA model suggested that lipophilicity, electron-richness, and lower degree of branching of the organic compounds are responsible for earthworm toxicity through various mechanisms. On the other hand, polar and bulky diverse chemicals do not have such toxic effects on earthworm. Hence, this model can be an effective tool to tailor molecular structures of the existing pesticides to develop novel compounds or pesticides which would be less toxic to the non-targeted organisms, specifically earthworm.

d. Regression and classification-based models on acute contact toxicity of organic binary mixtures in honey bees (*A. mellifera*)

This model has been described in the previous report, and in 2020 we published the paper regarding this model, containing the full details of the model.

e. Integrative QSAR model to predict acute contact toxicity (LD50) and to profile the Mode of Action (MoA) of pesticides active substances in honey bees (*A. mellifera*)

Pesticide toxicity data for honey bees (Apis mellifera) expressed as LD_{50} µg/bee (acute contact, 48h) were retrieved in June 2018 from EFSA's chemical hazards database "OpenFoodTox", US-EPA ECOTOX and Pesticide Properties DataBase to predict acute contact toxicity (LD_{50}) and to profile the Mode of Action (MoA) of pesticides active sub-stances.

Three different classification-based and four regression-based models were developed and tested for their performance, thus identifying two models providing the most reliable predictions based on k-NN algorithm.

Two different procedures have been adopted whether it was a classification or regression dataset building:

- **Classification-based models**
  The threshold used for toxicity classification was 100µg/bee, which corresponds to the limit test (OECD, 1998). If the values associated to the same SMILES fell under and up this threshold, the relative compound have been excluded for classification modelling. The final dataset is constituted by 413 compounds.
- **Regression-based model**
  All compounds presenting the qualifier (N) were excluded. All values were converted in µmol/bee and grouped by SMILES; geometric means for each value associated with the same SMILES were calculated. When values associated with the same SMILES showed a 3-fold difference between the maximum and the minimum, the relative compound was excluded from the regression modelling. When continuous data were available, these were trans-formed on the logarithmic scale. In addition, compounds excluded from the classification modelling were also excluded from the regression modelling. The final dataset was built from 113 compounds.

The two-category QSAR model (toxic/non-toxic; n= 411) was validated using sensitivity (=0.93), specificity (=0.85), balanced accuracy (=0.90), and Matthews correlation coefficient (MCC = 0.78) as statistical parameters. The regression-based model (n= 113) was validated for its reliability and robustness (R2= 0.74; MAE = 0.52). Current study proposes the MoA profiling for 113 pesticides active substances and the first harmonised MoA classification scheme for acute contact toxicity in honey bees, including LD50s data points from three different data-bases. The classification allows to further define MoAs and the target site of PPPs active substances, thus enabling regulators and scientists to refine chemical grouping and toxicity extrapolations for single chemicals and component-based mixture risk assessment of multiple chemicals. Relevant future perspectives are briefly addressed to integrate MoA,

adverse outcome pathways (AOPs) and toxicokinetic information for the refinement of single-chemical/combined toxicity predictions and risk estimates at different levels of biological organization in the bee health context.

Both datasets for classification-based and regression-based models were divided into a Training (TS) and Validation Set (VS). The number of compounds in each set is shown in Table 11.

Table 11: Datasets splitting for classification- and regression-based models.

	Classification – based models	Regression - based models
Train	328	88
Test	83	25
Tot	411	113

Table 12: Results of the statistical quality for random forest (RF), decision tree (DT) and k-nearest neighbor (k-NN) classification-based models. Test set (TS), cross validation set (CV) and validation set (VS) are reported.

Algorithm	RF			DT			K_NN		
Set	TS	CV	VS	TS	CV	VS	TS	CV	VS
Sensitivity	1.00	0.94	0.96	0.95	0.92	0.89	0.77	0.90	0.93
Specificity	1.00	0.54	0.74	0.58	0.53	0.89	0.96	0.67	0.85
Accuracy	1.00	0.81	0.89	0.83	0.80	0.89	0.90	0.83	0.90
MCC	1.00	0.54	0.75	0.59	0.50	0.76	0.76	0.59	0.78

Table 13: Statistical robustness and performance for random forest (RF), multi linear regression (MLR), partial least squares (PLS), decision tree (DT) and k-nearest neighbors (k-NN) regression-based models. Test Set (TS), Cross Validation set (CV) and Validation Set (VS) are reported.

Parameter	**RF**	**MLR**	**PLS**	**DT**	**KNN**	**Acceptability criteria**
RMSE - TS	0.41	0.66	0.76	0.71	0.39	
$r^{2`}$ TS	0.88	0.7	0.61	0.66	0.9	0.6
MAE - TS	0.34	0.56	0.61	0.53	0.3	

Parameter	RF	MLR	PLS	DT	KNN	Acceptability criteria
CCC - TS	0.93	0.82	0.76	0.79	0.94	
$Q^2_{5\text{-fold CV}}$	0.49	0.59	0.53	0.35	0.63	0.5
RMSE - VS	0.8	0.93	0.9	1.01	0.71	
r^2 - VS	0.72	0.55	0.59	0.46	0.74	0.6
MAE - VS	0.63	0.76	0.66	0.79	0.52	
Q2-F1	0.65	0.53	0.56	0.44	0.72	
Q2-F2	0.65	0.52	0.56	0.44	0.72	
Q2-F3	0.57	0.41	0.45	0.31	0.65	
CCC - VS	0.75	0.73	0.71	0.66	0.83	
r^2_0 - VS	0.65	0.53	0.58	0.46	0.73	
r^2_m - VS	0.53	0.48	0.52	0.43	0.66	0.5
$\bar{r}_2$ - VS	0.35	0.42	0.35	0.31	0.54	0.5
Δr^2_m - VS	0.36	0.11	0.34	0.23	0.23	<0.3
k - VS	0.98	0.96	0.93	0.94	0.96	0.85 < k < 1.15
k' - VS	0.93	0.91	0.95	0.91	0.97	0.85< k' < 1.15

f. Regression-based models for biocides - *Daphnia magna*

This model has been described in the previous report. In 2020 we published the paper regarding this model, containing the full details of the model.

g. Regression-based models for pesticides – *Daphnia magna*

There are many QSAR models developed on generic organic compounds, but they perform poorly when they are applied to pesticides. Thus, the development of QSAR models for toxicity of pesticides towards Daphnia magna remains the important task of QSAR analysis.

The present model is built up from a dataset of 308 experimental acute toxicity data (expressed as pEC50) of plant protection products and biocides for Daphnia magna. These short-term experimental data were retrieved from two large collections (the Office of Pesticide Programs (OPP) Pesticides Ecotoxicity Database and the EFSA's chemical hazards' database (OpenFoodTox)).
The Monte Carlo technique is the basis to build up the present QSAR model. This approach is attractive to build up predictive models since instead of a large number of different molecular descriptors the corresponding model is based on solely one optimal descriptor calculated with SMILES and all necessary calculations can be done using the CORAL software. Furthermore, the approach provides the mechanistic interpretation (e.g. aromaticity and branching of carbon skeleton are promoters of increase for toxicity towards Daphnia magna in the case of the examined set of pesticides).
The statistical quality of these models of acute toxicity of pesticides towards Daphnia magna is good: the best model is characterized by n=103, R2=0.76, RMSE=0.91 (training set); n=53, R2=0.82, RMSE=0.87 (validation set). The main advantage of the models suggested in this work is that they are based on molecular structure represented by SMILES and numerical data on the endpoint: no information on physicochemical data, 3D geometry, and descriptors of quantum mechanics is necessary for the modelling process.

h. Regression- and classification-based model (bobwhite quail)

Two datasets on acute oral toxicity (LD_{50}) in bobwhite quail (*C. virginianus*) have been created in order to develop two different QSAR models in CORAL software as follows:

- Regression model (n= 74), experimental values are reported in (log) mg/kg/bw;

- Classification model (n= 225), threshold of 2000 mg/kg/bw. In this case, two classes have been assigned i.e. LD_{50} > 2000 (Non toxic); LD_{50} < 2000 (Toxic) according to OECD GD 223.

Table 14: Statistical robustness and performance of the model are presented.

	Active training set	**Passive training set**	**Calibration set**	**Validation set**
N	60	58	58	58
TP	11	11	8	6
TN	49	47	38	38
FP	0	0	6	6
FN	0	0	6	8
Sensitivity	1.0	1.0	0.5714	0.4286
Specificity	1.0	1.0	0.8636	0.8636
Accuracy	1.0	1.0	0.7931	0.7586
MCC	1.0	1.0	0.4351	0.3087

i. Regression- based model for inhalation toxicity as No Observed Adverse Effect Concentration (NOAEC)

The inhalation toxicity studies the adverse effects that occur as a result of inhalation of chemical or biological agents and dust. In fact, inhaled agents can cause a wide range of negative health effects which include both mild local phenomena (for example irritation or sensitizing phenomena) and systemic diseases. Inhalation toxicity plays a key role not only in the occupational risk evaluation to assess the hazard of airborne xenobiotics to which workers are exposed during manufacturing process, but also in environmental and therapeutic exposure.
The adopted approach is based on information on the molecular structure and do not require other additional more complex data (such as 3D-descriptors, descriptors of quantum mechanics) or information on physicochemical or biochemical endpoints as used in other published papers. This feature makes the model more user-friendly without compromising the performance of the model itself, which is superior to what is already present in the literature.

The statistical characteristics of the best model for negative logarithm of NOAEC (pNOAEC) are for training set n = 108, average r^2 = 0.52+0.62+0.76 / 3 = 0.63 and for validation external set n = 35, r^2 = 0.73.

j. Regression-based model to predict water solubility (WS) of pesticides

In environmental risk assessment, knowledge of the physicochemical properties of pesticides is considered a key for defining the environmental fate of these compounds. Physico-chemical properties influence the distribution, impact, and fate of parent compounds or metabolites in the environment. According to REACH (2006), water solubility has a very high impact on the behaviour of substances and can be used as a criterion for assessing their ecotoxicological effects, hence, WS is one of the key parameters for assessing environmental properties.

Table 15. The statistical quality of the best CORAL model (abbreviations: CLB, calibration set; VLD, validation set)

	CCC_{CLB}	Rm^2_{CLB}	IIC_{CLB}	R^2_{VLB}
TF1	0.8393	0.6375	0.6147	0.7738
TF2	0.9204	0.7829	0.9238	0.7752
TF3	0.9433	0.8548	0.9435	0.8676

k. In silico model to predict mode of action, with reference to pesticides

Based on the studies done on pesticides and bees, we addressed the effects associated to the mode of action (MoA). This information is useful for bees' toxicity, but it is also valuable independently, for multiple uses. Of particular interest can be the application to group pesticides, and other substances, according to their MoA. Indeed, the widely used approach to address the mixture toxicity of substances is through the dose addition/concentration addition paradigm, which refers to the common effect of two or more sustances presenting the same MoA.
In particular, a set of more than 40 MoA have been identified, as in Table 16.

Table 16. The MoA for pesticides at the basis of the *in silico* model.

Class	MoA
Carbamate_AChE	Acetylcholinesterase (AChE) inhibitors_AChE(-)
Carbamate_MIT	Inhibition of mitosis/microtubule organization
Chloroacetamide (V1)	Inhibition of very-long-chain fatty acid synthesis (VLCFAs)
Cyclohexanedione(DIMs)	Inhibition of acetyl-CoA carboxylase (ACCase)
Dinitroaniline	Inhibition of microtubule assembly
Dinitrophenol_1	Uncoupling (membrane disruption)
Dinitrophenol_2	Uncoupling (membrane disruption)
Hydrazine carboxylate	Mitochondrial complex III electron transport inhibitors
Imidazole_1	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Imidazole_2	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Imidazole_3	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Imidazole_4	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Morpholine/spiroketal-amines_1	SBI: Class II_Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)
Morpholine/spiroketal-amines_2	SBI: Class II_Δ14-reductase and Δ8 to Δ7-isomerase in sterol biosynthesis (erg24, erg2)
Neonicotinoid	Nicotinic acetylcholine receptor (nAChR) competitive modulators_nACh-R(+)
Organochlorine_1	GABA-gated chloride channel blockers_GABA-R(-)
Organochlorine_2	Sodium channel modulators_Na channel(+)
Organochlorine/cyclodiene organochlorine_1	GABA-gated chloride channel blockers_GABA-R(-)
Organochlorine/cyclodiene organochlorine_2	GABA-gated chloride channel blockers_GABA-R(-)
Organophosphate	Acetylcholinesterase (AChE) inhibitors_AChE(-)
Oxadiazine	Voltage-dependent sodium channel blocker_Na channel(-)
Phenoxy-acetic acids	Synthetic auxins (action like indole acetic acid)
Pyridine carboxylic acid	Synthetic auxins (action like indole acetic acid)
Benzoic acid	Synthetic auxins (action like indole acetic acid)
Quinolone carboxylic acid & aromatic acetates	Synthetic auxins (action like indole acetic acid)
Phenylpyrazole(azole)	GABA-gated chloride channel blockers_GABA-R(-)
Phosphorodithionate	Inhibition of lipid synthesis – not ACCase
Pyrazolium(azole)	Mitochondrial complex I electron transport inhibitors
Pyrethroid/Plant derived/Pyrethrin	Sodium channel modulators_Na channel(+)
Pyrethroid/Synthetic_1	Sodium channel modulators_Na channel(+)
Pyrethroid/Synthetic_2	Sodium channel modulators_Na channel(+)
Pyridazinone	Mitochondrial complex I electron transport inhibitors
Pyrimidine	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Pyrrole(azole)	Uncoupler of oxidative phosphorylation
Sulfonylurea	Inhibition of acetolactate synthase (ALS)/acetohydroxyacid synthase (AHAS)
Sulfoximine	Nicotinic acetylcholine receptor (nAChR) competitive modulators
Sulphamide/electrophiles	Multi-site contact activity

Thiocarbamate	Inhibition of lipid synthesis – not ACCase
Triazine	Inhibition of photosynthesis at PS II
Triazole	SBI: Class I_C14-demethylase in sterol biosynthesis (erg11/cyp51)
Urea/amide	Inhibition of photosynthesis at PS II

These MoA have been codified into SMARTS format, and then a classifier model has been obtained, the *in silico* model has been implemented into VEGA to allow for its broad use.

4. Conclusions

This external scientific report provides an account of the maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. In particular, this document summarises the continued progress in the second year of the framework contract on OpenFoodTox 2.0. The activities proceeded regularly and according to plan, and cover both the maintenance and the new developments implemented in the second year. This is related both to the data and the *in silico* models developed. The OpenFoodTox database has been maintained, and new data gathered since the beginning of the project, also covering the temporal gap between the end of the previous OpenFoodTox related project (for its development) and the present one, has been added.

Further activities related to the present project include work done to adapt the OpenFoodTox database with the new data structure in order to be compliant to the new format of data organized within EFSA.

Another objective related to the project addresses the conceptual framework to host and exploit data of a heterogeneous nature, such as toxicokinetics. The current report documents the recent updates to the OpenFoodTox data model for physico-chemical properties and PK/TK data. In particular, the data models have been implemented using a custom XML format which has the necessary versatility to accommodate data values and types gathered from a variety of different sources and specifications. Physico-chemical property data has been differentiated into experimental and calculated types within the model schema, and the latter model can now retain details such as the software used for their computation. Finally, the PK/TK and ADE data models were further expanded to contain pertinent information at both the study and test levels.

OpenFoodTox content was enriched with physico-chemical properties on 183 substances and PK/TK data on 126 substances obtained from EFSA documents, as well as with quantitative values for key toxicokinetics parameters (Cmax, AUC, T-half and T-max) collected at the tested dose-level detail for 60 drugs from US FDA NDAs. Additionally, data gaps were also identified and annotated within the database.

A significant amount of work has been been devoted to the identification of suitable fields and cells, and to be organized within existing or redefined OHTs. This activity involved a series of fruitful meetings and collaborative work, extending beyond EFSA and the OptiTox consortium, including representatives of other European institutes such as the JRC, OECD representatives have also been contacted and contributed to the discussion.

To this end, data models were developed for physicochemical properties and PK/TK data by designing a reusable, custom XML format capable of recording various types of property values and conditions, as well as annotations for reported values. In the second year of the project, an updated version of the XML structure was been created.

Finally, beyond simple maintainance activity, a larger framework has been identified and the new OpenFoodTox database will benefit through the added ability to gather data related to the toxicokinetic information derived from the dossiers.

Furthermore, the existing OpenFoodTox data can be linked with other data of varying nature, such as exposure values.

The exploitation of the new, extended OpenFoodTox database, expanded with multiple properties as mentioned above, will proceed to the development and optimization of *in silico* models. These models will provide an initial value to be used in particularly data-poor scenarios. For this purpose, studies have been conducted on the integration of results within a WoE approach, and it is likely that multiple *in silico* models and read-across tools exist for the same endpoint.

In this year, a number of *in silico* models have been studied and developed, as documented in published papers, with several others in the pipeline. Furthermore, papers introducing structure, novelty and

future developments of OpenFoodTox have been published as a result of our efforts (Dorne et al., 2021).

The activities done within this year illustrate that it is feasible to cope with the ambitious objectives of OptiTox, that is, the development of the advanced version of OpenFoodTox which integrates multiple lines of evidence for better risk assessments. The new OpenFoodTox will represent a pillar within the framework of EFSA activities, and at the same time will establish links with other regulatory authorities both inside and outside of Europe. This methodological and technological improvement is expected to re-shape the assessment of substances, with a deeper understanding of the network of processes provoking the final toxicological outcome. Networking with other software systems, such as VEGA (www.vegahib.eu) is anticipated, in order to fill data gaps and support the evaluation of substances using predicted values and it will be addressed starting from next year. Similarly, further data on relevant proteries will be included (e.g. bioaccumulation and metabolism). In addition, further work on OHT data format and the most suitable way to reporting the combination of WoE biological relevance and uncertainty in envisaged in the next year.

References

Alla P. Toropova, Andrey A. Toropov, Edoardo Carnesecchi, Emilio Benfenati, Jean Lou Dorne, The using of the Index of Ideality of Correlation (IIC) to improve predictive potential of models of water solubility for pesticides. Environmental Science and Pollution Research (2020) 27: 13339–13347.

Andrey A. Toropov, Alla P. Toropova, Emilio Benfenati, QSAR model for pesticides toxicity to Rainbow Trout based on "ideal correlations". Aquatic Toxicology 227 (2020) 105589.

Andrey A. Toropov, Alla P. Toropova, Diego Baderna, Gianluca Selvestrel, Emilio Benfenati, *in press*, Prediction of No Observed Adverse Effect Concentration for Inhalation toxicity using Monte Carlo approach. SAR and QSAR in Environmental Research

Benfenati E, Carnesecchi E, Roncaglioni A, Baldin R, Ceriani L, Ciacci A, Kovarich S, Sartori L, Mostrag A, Magdziarz T, Yang C, 2020. Maintenance, update and further development of EFSA's Chemical Hazards: Open-FoodTox 2.0. EFSA supporting publication 2020:EN-1822. 36 pp. doi:10.2903/sp.efsa.2020.EN-1822

Emilio Benfenati, Qasim Chaudhry, Giuseppina Gini, Jean Lou Dorne, 2019. Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy, Environment International, Volume 131, 2019, 105060, ISSN 0160-4120, https://doi.org/10.1016/j.envint.2019.105060

Cappelli CI, Toropov AA, Toropova A, Benfenati E. Ecosystem ecology: Models for acute toxicity of pesticides towards Daphnia magna. Environmental Toxicology and Pharmacology, 2020, 80, 103459. https://doi.org/10.1016/j.etap.2020.103459.

Carnesecchi E, Toropov AA, Toropova AP, Kramer N, Svendsen C, Dorne JL, Benfenati E., 2020. Predicting acute contact toxicity of organic binary mixtures in honey bees (A. mellifera) through innovative QSAR models. Science of the Total Environment, 2020, 704, 135302, https://doi.org/10.1016/j.scitotenv.2019.135302

Carnesecchi E, Toma C, Roncaglioni A, Kramer N, Benfenati E, Dorne JLCM, 2020a. 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, Volume 735, 15 September 2020, 139243. https://doi.org/10.1016/j.scitotenv.2020.139243.

Dorne JLCM, Richardson J, Livaniou A, Carnesecchi E, Ceriani L, Baldin R, Kovarich S, Pavan M, Saouter E, Biganzoli F, Pasinato L, Zare Jeddi M, Robinson TP, Kass GEN, Liem AKD, Toropov AA, Toropova AP, Yang C, Tarkhov A, Georgiadis N, Di Nicola MR, Mostrag A, Verhagen H, Roncaglioni A, Benfenati E, Bassan A, 2021. EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments. Environment International 146 (2021) 106293.

European Commission (EC), 2000. Commission Regulation No 1565/2000 of 18 July 2000 laying down the measures necessary for the adoption of an evaluation programme in application of Regulation (EC) No 2232/96. Official Journal of the European Communities 19.7.2000, L 180, 8-16.

EFSA (European Food Safety Authority), 2005a. Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. The EFSA Journal (2005) 282, 1-31.

EFSA (European Food Safety Authority), 2006a. Tolerable upper intake levels for vitamins and minerals. Scientific Committee on Food Scientific Panel on Dietetic Products, Nutrition and Allergies. February 2006. www.efsa.europa.eu/en/ndatopics/docs/ndatolerableuil.pdf

EFSA (European Food Safety Authority), 2008a. Polycyclic Aromatic Hydrocarbons in Food, The EFSA Journal (2008) 724, 1-114.

EFSA (European Food Safety Authority), 2009a. European Food Safety Authority; Guidance Document on Risk Assessment for Birds & Mammals on request from EFSA. EFSA Journal 2009; 7(12):1438. doi:10.2903/j.efsa.2009.1438.

EFSA (European Food Safety Authority), 2009b. Marine biotoxins in shellfish – Saxitoxin group. The EFSA Journal (2009) 1019, 1-76.

EFSA (European Food Safety Authority), 2012a. Completion of data entry of pesticide ecotoxicology Tier 1 study endpoints in a XML schema – database. Pierobon E., Neri M.C., Marroncelli S., Croce V.; Supporting Publications 2012:EN-326. [96 pp.].

EFSA (European Food Safety Authority), 2012b. Guidance for submission for food additive evaluations. EFSA Journal 2012;10(7):2760.

EFSA (European Food Safety Authority), 2012c. Scientific Opinion on Exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). EFSA Journal 2012;10(7):2750.

EFSA (European Food Safety Authority), 2013a. Report on 'Data collection and data entry for EFSA's chemical hazards database NP/EFSA/EMRISK/2011/01', S-IN Soluzioni-Informatiche; Supporting Publications 2013:EN-458. [140 pp.]. Available online: http://www.efsa.europa.eu/en/supporting/pub/458e.htm

EFSA (European Food Safety Authority), 2013b, Standard Sample Description ver. 2.0, EFSA Journal 2013;11(10):3424 [114 pp.]. Available online: http://www.efsa.europa.eu/en/efsajournal/pub/3424.htm

EFSA (European Food Safety Authority), 2014. Report on 'Further development and update of EFSA's Chemical Hazards Database NP/EFSA/EMRISK/2012/01', S-IN Soluzioni-Informatiche; Supporting Publications 2014: EN-654. [103 pp.]. Available online: http://www.efsa.europa.eu/it/supporting/pub/654e.htm

EFSA (European Food Safety Authority), 2015. Further development and update of EFSA's Chemical Hazards Database. EFSA supporting publication 2015: EN-823. [84 pp.]. Available online: http://www.efsa.europa.eu/it/supporting/pub/823e

EFSA (European Food Safety Authority), 2017. Update and maintenance of EFSA's Chemical Hazards Database. EFSA supporting publication 2017: EN-1192. [73 pp.]. Available online:

https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2017.EN-1192EPA (Environmental Protection Agency) 1993. Reference Dose (RfD): Description and Use in Health Risk Assessments Background Document 1A March 15, 1993. http://www.epa.gov/iris/rfd.htm

EPA (Environmental Protection Agency), 2011. Vocabulary Catalog List Detail - Integrated Risk Information System (IRIS) Glossary. http://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do?details=&glossaryName=IRIS%20Glossary, last updated: August 31, 2011.

FDA (Food and Drug Administration), 2005. Guidance for Industry Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER), 2005.

Ghosh S, Ojha P K, Carnesecchi E, Lombardo A, Roy K, Benfenati E Exploring QSAR modeling of toxicity of chemicals on earthworm. Ecotoxicol Environ Saf 2020; 190: 110067 IF: 4.872

Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V, Jaeschke R, Rind D, Dahm P, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Murad MH, Schünemann HJ; GRADE Working Group. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011 Dec;64(12):1311-6. doi: 10.1016/j.jclinepi.2011.06.004. Epub 2011 Jul 30. PMID: 21802902.

ILSI (International Life Sciences Institute), 2009. Application of the Margin of Exposure Approach to Compounds in Food which are both Genotoxic and Carcinogenic. Summary Report of a Workshop held in October 2008 Organised by the ILSI Europe Risk Assessment of Genotoxic Carcinogens in Food Task Force. Pushing date: October 2009.

IPCS/OECD (International Programme on Chemical Safety/ Organisation for Economic Cooperation and Development), 2004. Harmonization Project Document No. 1. IPCS Risk Assessment Terminology. World Health Organization Geneva, 2004.

Kroes R, Renwick A G, Cheeseman M A, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos J G, Wurtzen G, 2004. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food and Chemical Toxicology 42: 65-83.

Li, Y. F., and Ngom, A. (2015). "Data integration in machine learning" in Proceedings 2015 IEEE International Conference on Bioinformatics and Biomedicine (Washington, DC), 1665-1671

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. Sci Total Environ. 2009 Sep 15;407(19):5199-205. doi: 10.1016/j.scitotenv.2009.05.004. Epub 2009 Jul 19. PMID: 19619890.

M. Marzo, G.J. Lavado, F. Como, A.A. Toropov, A.P. Toropova, D. Baderna, C. Cappelli, A. Lombardo, C. Toma, M. Blázquez Sánchez, E. Benfenati, QSAR models for Biocides. The example of the prediction of Daphnia Magna acute toxicity. SAR and QSAR in Environmental Research, 31:3, 227-243, 2020.

MeSH (Medical Subject Headings), 2014a. NLM controlled subject vocabulary used for indexing and cataloging: https://www.nlm.nih.gov/mesh/

MeSH (Medical Subject Headings), 2014b.NLM controlled subject vocabulary used for indexing and cataloging. Files available for download: https://www.nlm.nih.gov/mesh/filelist.html ftp://nlmpubs.nlm.nih.gov/online/mesh/.xmlmesh/desc2014.zip (Descriptor Records XML file),

ftp://nlmpubs.nlm.nih.gov/online/mesh/.xmlmesh/supp2014.zip (Supplementary Concept Records XML file).

National Toxicology Program (NTP) 2015. Handbook for Preparing Report on Carcinogens Monographs. National Toxicology Program, Office of the Report on Carcinogens. http://ntp.niehs.nih.gov/pubhealth/roc/handbook/index.html.

OECD (Organisation for Economic Co-operation and Development), 1998 Test No. 214: Honeybees, Acute Contact Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2 OECD Publishing, Paris (1998), 10.1787/9789264070189-en

OECD (Organisation for Economic Cooperation and Development), 2012. Picklist for the OECD harmonized templates. (Version 3, February 2012).

Office of Health Assessment and Translation (OHAT). 2015. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Toxicology Program

OSHA European Agency for Safety and Health at Work, 2016. Annual Report 2015 https://osha.europa.eu/sites/default/files/publications/documents/TEAF16001ENN_AR15-summary.pdf

Renwick AG, 2005. Structure-based thresholds of toxicological concern – guidance for application to substances present at low levels in the diet. Toxicology and Applied Pharmacology 207 (Supplement): S585- S591.

Roy J, Ojha P K, Carnesecchi E, Lombardo A, Roy K, Benfenati E First report on a classification-based QSAR model for chemical toxicity to earthworm J Hazard Mater 2020; 386: 121660

SCF (Scientific Committee on Food), 1999. Opinion on a programme for the evaluation of flavouring substances (expressed on 2 December 1999). Scientific Committee on Food. SCF/CS/FLAV/TASK/11 Final 6/12/1999. Annex I the minutes of the 119th Plenary meeting. European Commission, Health & Consumer Protection Directorate-General.

S-IN Soluzioni Informatiche, 2018. Final report on the update and maintenance of EFSA's Chemical Hazards Database. Submitted to EFSA on 6th April 2018, currently under revision.

van Leeuwen CJ, 2007. Risk Assessment of Chemicals: An Introduction. Springer.com

Verhaar HJM, van Leeuwen CJ and Hermens JLM (1992) Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. Chemosphere 25, 471-491.

World Cancer Research Fund/American Institute for Cancer Research Second Expert Report – Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective WCRF/AICR, Washington (2007).

Appendix A – User Manual of the EFSA's Chemical Hazards Database

Appendix B – Dihydroeugenol intermediate effects (1)

Appendix C – Dihydroeugenol intermediate effects (2)

Appendix D – Fipronil intermediate effects

The appendices are provided as additional files.