Regulatory Toxicology and Pharmacology 142 (2023) 105416

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

A non-clinical and clinical IUCLID database for 530 pharmaceuticals (part I): Methodological aspects of its development

Martina Evangelisti [a,1], Marco Daniele Parenti [a,c,1], Greta Varchi [c], Jorge Franco [a,d], Jochen vom Brocke [b], Panagiotis G. Karamertzanis [b], Alberto Del Rio [a,c,*], Ingo Bichlmaier [b,**]

[a] Innovamol Consulting Srl, Via San Faustino 167, 41126, Modena, Italy
[b] European Chemicals Agency (ECHA), Telakkakatu 6, 00150, Helsinki, Finland
[c] Institute for Organic Synthesis and Photoreactivity (ISOF), National Research Council of Italy (CNR), Via P. Gobetti 101, I-40129 Bologna, Italy
[d] Genoa University, Via Balbi 5, 16126, Genoa, Italy

ARTICLE INFO

Handling Editor: Dr. Daniele Wikoff

Keywords:
IUCLID database
Animal testing
Human information
Standard product label
Ontology
Effect levels
Endocrine disruption
Repeat-dose toxicity
Carcinogenicity
Reproductive toxicity
Developmental toxicity
Animal-human correlation
New approach methodologies
NAMs

ABSTRACT

A new IUCLID database is provided containing results from non-clinical animal studies and human information for 530 approved drugs. The database was developed by extracting data from pharmacological reviews of repeat-dose, carcinogenicity, developmental, and reproductive toxicity studies. In the database, observed and no-observed effects are linked to the respective effect levels, including information on severity/incidence and transiency/reversibility. It also includes some information on effects in humans, that were extracted from relevant sections of standard product labels of the approved drugs. The database is complemented with a specific ontology for reporting effects that was developed as an improved version of the Ontology Lookup Service's mammalian and human phenotype ontologies and includes different hierarchical levels. The developed ontology contains novel and unique standardized terms, including ontological terms for reproductive and endocrine effects. The database aims to facilitate correlation and concordance analyses based on the link between observed and no-observed effects and their respective effect levels. In addition, it offers a robust dataset on drug information for the pharmaceutical industry and research. The reported ontology supports the analyses of toxicological information, especially for reproductive and endocrine endpoints and can be used to encode legacy data or develop additional ontologies. The new database and ontology can be used to support the development of alternative non-animal approaches, to elucidate mechanisms of toxicity, and to analyse human relevance. The new IUCLID database is provided free of charge at https://iuclid6.echa.europa.eu/us-fda-toxicity-data.

1. Introduction

Over the last two decades, information on chemicals has evolved rapidly due to the need to collect, organise, and retrieve information from chemical data for research and regulatory purposes. In the context of regulating chemicals, the Regulation on Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH, (EC) No 1907/2006) (European Parliament and Council of the European Union, 2006) was adopted to improve the protection of human health and the environment from the risks posed by industrial chemicals. The regulation embeds the need of collecting chemical information in domains not only related to industrial processes but also our day-to-day lives, for example cleaning products, paints, clothes, furniture, and electrical appliances. Many other pieces of legislation rely on the information generated under REACH, for example, the Classification, Labelling and Packaging (CLP) Regulation ((EC) No 1272/2008) (European Parliament, 2008).

To support the collection of chemicals information, the International Uniform Chemical Information Database (IUCLID) was created as a comprehensive database on chemicals as a result of the European Commission's need to assess the risks of chemicals placed on the European market before September 18, 1981 (Council, 1993). IUCLID contains information on environmental and toxicological endpoints, as well as use and exposure data. The database is used to evaluate hazard information of registered substances and to identify substances of concern.

* Corresponding author. Innovamol Consulting Srl, Strada San Faustino 167, 41126, Modena, Italy.
** Corresponding author. European Chemicals Agency (ECHA), Telakkakatu 6, 00150, Helsinki, Finland.
E-mail addresses: alberto.delrio@innovamol.com (A. Del Rio), ingo.bichlmaier@echa.europa.eu (I. Bichlmaier).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.yrtph.2023.105416
Received 6 December 2022; Received in revised form 4 May 2023; Accepted 21 May 2023
Available online 28 May 2023

In 2000, IUCLID also started to include active ingredients in biocidal products with the objective of creating a list of approved active ingredients based on substance data (Heidorn et al., 2003; European Parliament and Council, 2012). The European Chemicals Agency (ECHA) co-develops the software with the Organization for Economic Cooperation and Development (OECD) (ECHA, 2022a).

On the other hand, the requirement for chemical substance registration, including toxicological information, has resulted in the development of tools and resources necessary for structuring hazard data and assessing chemical risks. Along with the aforementioned IUCLID, these include databases, ontologies, and dictionaries like MedDRA, ToxRefDB and others (Bodenreider; Knudsen et al., 2013; Cai et al., 2015; Watford et al., 2019). These tools can be used to create standard product information and documentation for marketing authorisation applications as well as to record adverse events for expedited submission of safety data to governmental regulatory authorities. However, toxicological data structuring still has several difficulties. Among them are poor or inaccurate data quality, a lack of uniform data standardisation, the requirement for ongoing data updating, the restriction of data access due to concerns about confidentiality or intellectual property, and various classification interpretations that can result in inconsistency or misunderstanding.

For toxicological data to be used effectively in hazard and risk assessment, these obstacles can be overcome. Accordingly, IUCLID offers the possibility to store huge amounts of data that need to include good-quality information to be entered according to specific formats. For instance, REACH requires registrants to submit data using the OECD harmonised templates (OHTs) (ECHA, 2022b). This has led to the successful registration of over 23000 substances, of which approximately 4500 are produced in tonnages greater than 100 tons/year, therefore needing to meet Annexes IX or X requirements of the REACH regulation. In particular, *in vivo* animal studies on repeated-dose toxicity (RDT) to identify specific target organs, on carcinogenicity, on reproductive toxicity, i.e. sexual function and fertility, and on developmental toxicity need to be provided for those substances. Data generated from industrial chemicals, biocides, pesticides, or pharmaceuticals are equally relevant and complementary for scientific data analysis. Similarly, using larger datasets and applying the training set to substances with wider structural diversity, as for data pooled from different regulatory regimes, greatly benefit the development of predictive models.

One important aspect is the lack of mechanistic information on humans, for substances registered under REACH. Since such information is usually unavailable for industrial chemicals, certain data analyses are not straightforward, for example, concordance analyses between non-clinical (animal) and clinical (human-derived) data. On the other hand, substances that are used for pharmaceutical purposes allow to collect useful mechanistic and clinical information. Another advantage is that the regulatory requirements on animal studies on pharmaceuticals largely address the same study types or endpoints as *in vivo* animal toxicity studies for industrial chemicals under REACH. These studies, as mentioned above, include RDT, carcinogenicity, reproductive toxicity (sexual function and fertility) and developmental toxicity.

Non-clinical animal data were obtained in an unstructured format from the pharmacological reviews that are published by the US Food and Drug Administration (US FDA) (Drugs@FDA, 2022). Human information can be obtained from standard product label (SPL) files that are also provided by the US FDA (Drugs@FDA, 2022). Despite the free availability of these files, their usefulness for statistical analysis and model development is limited due to their unstructured nature, particularly when it comes to observed effects that lack an appropriate ontology. Interestingly, in recent years, there has been a growing interest in the development of ontologies covering toxic effects of chemical substances at various levels of organization. This trend is demonstrated by a range of efforts, such as the study by Ives et al. which provides an overview of the challenges and opportunities in developing ontologies for chemical risk assessment and management (Ives et al., 2017). Another example is the Environmental Health Language Collaboration, which aims to harmonize terminology and data representation across different environmental health disciplines (Holmgren et al., 2021). Specific efforts have also been made to ontologize IUCLID data, as exemplified by the ontologies available through the QSAR Toolbox (QSAR Toolbox, 2023). These approaches underline the importance of establishing a unified and standardized approach for toxicological information, which can facilitate better communication, integration, and analysis across various domains and applications. In fact, the lack of standardisation makes it difficult, if not impossible, to correlate information of pharmacological reviews and SPL data. However, animal data and human information can be used to support the evaluation of effects exerted by industrial chemicals, to help understand the mechanistic bases of their toxicity and to derive associations between animals and humans. For instance, an interesting assumption is that the observed correlations between animal and human data for pharmaceuticals can also apply to industrial chemicals since, toxicity pathways, adverse outcome pathways (AOP) and mode-of-action (MoA), which are causal chains of biochemical and biological events, can typically start from any initiating molecule leading to an adverse effect. Therefore, for the same principle, pharmaceutical data could also be useful for further validating the test guidelines and identifying chemistries and effects for which the relevance of the test results to humans is questionable.

This work describes the generation of IUCLID datasets and a new ontology for animal studies and human information of approved pharmaceuticals. The IUCLID datasets were compiled by systematically transferring toxicity data from the pharmacological reviews and SPLs into the OECD harmonised templates as encoded in IUCLID and extended with an ontology-based description of observed effects. 530 new drug applications (NDAs) were analysed to extract observed and no-observed effects and their effect levels as reported in RDT, carcinogenicity, reproductive toxicity and developmental toxicity studies and for SPLs warnings, precautions and adverse reactions data. Among others, IUCLID can be used to perform correlation and concordance analyses between animal and human data, which can help identify human-relevant endpoints and promote the development of alternative non-animal approaches for drug safety and efficacy evaluation. It also aims to provide structured information to increase scientific and regulatory knowledge, provide useful data for (quantitative) structure-activity relationships ((Q)SAR) or mechanistic studies. The increased toxicity knowledge and the development of mechanistic and predictive models further promote safe-by-design approaches (Knight et al., 2021) and has significant value for pharmaceutical industries, regulatory authorities, and researchers in advancing drug development while reducing the reliance on animal testing.

2. Materials and methods

2.1. Original data source and nature of the data to be structured

The Drugs@FDA database (Drugs@FDA, 2022) contains information on pharmaceutical substances submitted for market authorisation to the US. For this work, new drug applications (NDAs) for new molecular entities (NMEs) were selected because their files contain specific information on animal and clinical studies. Specifically, the NDAs used in this work relate to information on new drugs for which safety and effectiveness have been shown to meet regulatory requirements for marketing approval by the US Food and Drug Administration (US FDA).

From the data point of view, an NDA dossier must contain information on chemistry, pharmacology, medicine, biopharmaceutics and statistics. There are several categories of NDAs for pharmaceutical products depending on whether the product contains an NME (the so-called type 1 category NDA, thus containing the active moiety), new active ingredients (when a salt, ester or non-covalent derivative is added to the active moiety), new dosage forms, or formulations (Center for Drug Evaluation and Research, 2015). Herein, mainly type 1 NDAs were

considered since their dossiers contain pharmacological reviews and standard product label (SPL) files representing the onset of this work. The NDAs were further filtered to focus on active pharmaceutical ingredients (APIs) representing molecules that were relatively small, moderately lipophilic, *i.e.* generally adhering to the Lipinsky rule of five, known to have a mechanism of action (MoA) and that are generally used in single-drug formulations. Amino acids, peptides, oligonucleotides, antibodies, sugars, inorganic substances, and contrast agents were also processed.

Given that the project could not structure the whole NDA database, an initial analysis was performed as a sampling procedure to allow selecting NDAs based on a wide pharmacological diversity. We used the distribution of drugs as obtained from European Medicines Agency (EMA) data to replicate the pharmacological diversity in terms of selected NDAs. To achieve this, we used the Anatomical Therapeutic Chemical Classification System (ATC) of the World Health Organization (WHO), with special reference to anatomical class as selection parameter. The list of EMA drugs was downloaded from the EMA's website (EMA, 2018) and filtered for approved and non-generic drugs. Then, the anatomical group of each drug was used to calculate the distribution in the EU market and a similar distribution was obtained by appropriate selection of NDAs (DrugBank Online, 2022; FDA, 2019; Drugs.com, 2022), as depicted in Fig. 1.

The obtained final list of NDAs considered in this study is reported in the Supplementary Table S1.

2.2. Data sections in each new drug application (NDA)

For each NDA, the pharmacological review and the SPL were downloaded in pdf format from the US FDA website (FDA, 2022). As a starting point, the NDA number was searched from this website. In the "Approval Date(s) and History, Letters, Label, Review for NDA xxxxxx" section, all necessary documents for approval can be found. The Pharmacological Review document was downloaded from the "Original Approvals or Tentative Approvals" section. This document can either be directly attached as a pdf file or can be found under the "Review" section. In more recent NDAs, the document may also be labelled as "Multi-Discipline Review" or "Non-Clinical Review". Data usages of these files are generally free for both commercial and non-commercial application. Certain files included confidential data with specific redacted text. However, this confidential information did not impact the major objectives of this work.

Regarding pharmacological reviews, relevant data sections belonged to:

- RDT studies in which the drug was administered repeatedly (*e.g.,* once or twice a day) for a variable period (*e.g.,* four weeks, six months or one year). Generally, both male and female animals are exposed to the drug, and the study's objective is to detect the systemic toxicity and specific target organ toxicity. These studies include observations on mortality, clinical signs, body weight, food consumption, haematology, clinical biochemistry, organ weight, gross pathology, and non-neoplastic histopathology.
- Carcinogenicity studies, in which the drug is usually administered with an exposure duration of 104 weeks, are carried out in two species of male and female test animals. Mortality, and incidence of neoplastic formations in response to chronic exposure of substances are specifically analysed, emphasising histopathological changes.
- Reproductive toxicity studies (also referred to as segment I and segment III studies) typically investigate sexual function and fertility including parturition (live birth) to provide living pups, thus investigating pre-mating, mating, gestation, parturition and post-parturition (lactation). The effects of the drug on sexual function/fertility are evaluated for parental male and female animals, and for developmental aspects in pups.
- Developmental toxicity studies are usually referred to as segment II studies and the drug is tested in pregnant females during organogenesis, thus approximately between gestation day 6 and 19/20 in rats and gestation day 6 and 28/29 in rabbits, with the main objective to detect malformations and variations in foetuses after a caesarean section.

All pharmacological review studies providing information about the toxicity of interest were collected as originally reported by the medical writer, converted into text, and transferred into a single pdf document per study. No alteration was introduced unless required for optical character recognition (OCR) described below. Studies that contained no relevant effects were not collected.

Regarding SPLs, the label file was downloaded from the Drugs@FDA website, in the "Original Approvals or Tentative Approvals" section as PDF file. Sections were collected related to warnings, precautions, and adverse reactions. For warnings and precautions, the focus was on human information relating to carcinogenicity, mutagenicity, drug-drug interactions, and pregnancy categories reported as a descriptive text. Adverse reactions observed in human populations were collected with information on the relative incidence expressed as a percentage. It is worth noting that SPLs included results from clinical studies conducted before the approval of the drug and post-marketing surveillance data (pharmacovigilance). Only one SPL per NDA dossier was recorded.

For both documents, *i.e.* the Pharmacological Review and the Label

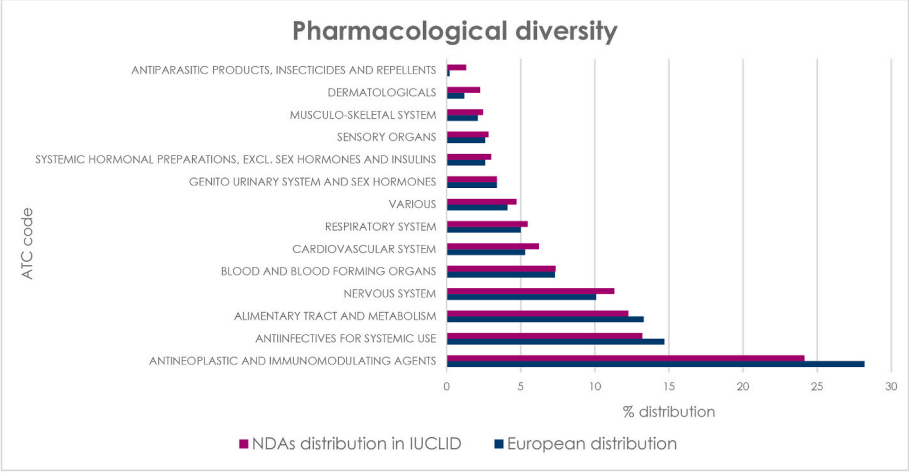

Fig. 1. Analysis showing the distribution of drugs in the European market by using the anatomical ATC code and the selected distribution of NDAs to be translated into IUCLID.

files, the date of access should be intended the date at which the PDF files were downloaded as listed in Table S1, while the data extraction has been performed subsequently.

2.3. Information persisted in IUCLID

2.3.1. Substance identity

General drug information included generic name, IUPAC name, code names, pharmacological class, CAS number, molecular formula, molecular weight, and structural formula of the NME. This section also included the composition of the drug formulation administered during the studies, *i.e.* any additional constituents, impurities or additives together with its purity. This information was later on included in the relevant studies in IUCLID, whilst the NME identifiers were added in the reference substance attached to the substance that is the dataset subject. Substance identity information, reference substance and test material definition were entered in IUCLID as depicted in the example of Tables S2–S4 of the supplementary information. The substance list contains the substance name and other identifiers and can be used to examine the overlap of the created database with other inventories, *e.g.* to see the amount of new information this work brings compared to pre-existing training sets or other toxicity databases.

2.3.2. Endpoint study records

Studies: For RDT, carcinogenicity, reproductive and developmental toxicity studies the following information was located and structured into IUCLID as detailed in Tables S2–S4 of the supplementary information:

- Materials and methods, *i.e.* endpoint, test material, species, strain, sex and related details like age and weight of animals.
- Administration and exposure, *i.e.* route of administration, vehicle, dose volume, analytical verification of doses, duration of treatment and its frequency, number of animals per sex per dose, details of the study design and doses/concentrations.
- Effects, effect levels and target systems *i.e.* effect and effect levels, target system and organ toxicity.
- Examination results, *i.e.* relevant text reported by the medical writer divided in sections, for instance, clinical signs, mortality, food consumption, haematological findings, histopathological findings and so forth.
- Other information, *i.e.* conclusion and executive summary of the study

Specific study information was included in the following harmonised templates in IUCLID.

- Repeated dose toxicity - Oral: OECD Harmonised Template N° 67
- Repeated dose toxicity - Inhalation: OECD Harmonised Template N° 68
- Repeated dose toxicity - Dermal: OECD Harmonised Template N° 69-1
- Repeated dose toxicity - Other Routes: OECD Harmonised Template N° 69-2
- Carcinogenicity: OECD Harmonised Template N° 72
- Reproductive toxicity: OECD Harmonised Template N° 73
- Developmental toxicity: OECD Harmonised Template N° 74

Standard Product Labels: Regarding SPL, the focus was on the warnings, precautions and adverse reactions with their incidence expressed as a percentage. SPL information was included in the harmonised templates OHT 81 (Direct observations: clinical cases, poisoning incidents and others) in IUCLID, in particular, in the fields depicted in Table S5.

2.3.3. Conventions used to create endpoint study records

In order to structure doses, concentrations, effects and effect levels, the annotation of each study record and standard product label (SPL) was performed by following a standard operating procedure (SOP) that was developed in an iterative manner. The SOP considered all the information blocks mentioned in the method section. It is noteworthy that a lot of information is constituted essentially by text that could be used as such in IUCLID, for instance: study type and design, the adequacy of study, animal species and strain, sex, age, route of administration, vehicle, duration of the treatment and frequency and number of animals used per sex per dose. On the other hand, this approach was not sufficient for doses/concentrations, effects and effect levels That required establishing conventions and rules to ensure consistent data entry. Table 1 summarises the rules for reporting doses and concentrations.

A comprehensive set of rules was also applied to report effect and effect levels as shown in Table 2. In general, only test-related effects and effect levels were reported while, for instance, effects and effect levels that were unrelated to the drug or not statistically significant were not reported in order to avoid the inclusion of unnecessary data in the database. Whenever available, the following details were reported.

i) degree of severity (*e.g.* slight, weak, moderate, marked, strong, severe)
ii) transient effect if the effect reduced/vanished while still under treatment (reported as “transient”)
iii) reversible effect if the effect reduced/vanished during the non-treatment recovery period (report as “reversible”).

Each effect and effect level were recorded with the basis for effect level (class of effect) from the picklist available in IUCLID, including:

- mortality
- clinical signs
- body weight and weight gain
- food consumption and compound intake
- food efficiency
- water consumption and compound intake
- ophthalmological examination
- haematology
- clinical biochemistry
- urinalysis
- behaviour (functional findings)
- immunology
- organ weights and organ/body weight ratios
- gross pathology
- neuropathology
- histopathology: non-neoplastic
- histopathology: neoplastic
- dermal irritation
- other

The no observed adverse effect level (NOAEL) was generally reported

Table 1
SOP to report doses for different observed cases.

Example of observed case	Way to report the effect level in IUCLID (mg/kg bw/day)
Control dose of 0 mg/kg/day	0
Low dose at 10 mg/kg/day	10
Medium dose 1 at 100 mg/kg/day	100
Medium dose 2 at 300 mg/kg/day	300
High dose at 1000 mg/kg/day	1000
Dose reduction at day 10, passing from 100 to 10 mg/kg/day - effect observed within the first 10 days	100
Dose reduction at day 10, passing from 100 to 10 mg/kg/day - effect observed after the 10th day	10
Dose termination at day 10, dose of 1000 mg/kg/day	1000

Table 2
SOP to report effects and effect levels considering different encountered cases in a typical study with low dose (LD) at 10 mg/kg/day, medium dose 1 (MD1) at 100 mg/kg/day, medium dose 2 (MD2) at 300 mg/kg/day and high dose (HD) at 1000 mg/kg/day.

Example of observed case	Dose picklist in IUCLID	Mathematical inequalities	Effect level as reported in IUCLID	Additional remarks included in IUCLID
No effect case	NOEL	≤	1000	
Low dose (*e.g.* 10 mg/kg/day)	Dose level	≥	10	*e.g.* slight, reversible
Medium dose 1 (*e.g.* 100 mg/kg/day)	Dose level	≥	100	*e.g.* slight, reversible
Medium dose 2 (*e.g.* 300 mg/kg/day)	Dose level	≥	300	*e.g.* slight, reversible
High dose (*e.g.* 1000 mg/kg/day)	Dose level	=	1000	*e.g.* slight, reversible
Effect in an interval case (*e.g.* between 10 and 1000 mg/kg/day)	Dose level	≥ <	10 1000	*e.g.* slight, reversible
NOAEL dose reported (*e.g.* 300 mg/kg/day)	NOAEL	=	300	
NOAEL dose not reported	NOAEL		N/A	"not reported by medical writer" or "not determined by medical writer"
No dose reported case	N/A		N/A	"no effect level reported"

as identified by the medical writer but, in general, a holistic approach was adopted to make explicit what exactly the pharmacological review contained even if not mentioned explicitly. This holistic approach was used to strive to maintain consistency in the database of all NDAs. If the NOAEL was not reported, meaning not explicitly mentioned by the medical writer, the convention was to add in the IUCLID field the specific comment "not reported" while if the medical writer mentioned the NOAEL without assigning the specific dose it was reported a specific comment "not determined". In the infrequent case where an interval was described, for instance in a sentence of this kind "Cumulative food consumption was significantly decreased in low- and mid-dose females compared to control values, but not in the high-dose females" the convention was indicated both starting and ending dose correspondent to the effect reported.

Other information regarding each study consisted mainly of text or picklist and was reported only when originally available in the pharmacological review. This information is described in the previous paragraph 2.3.2.

2.4. New ontology

2.4.1. Limitation of IUCLID OHTs and existing picklists

A current limitation of OHTs is that effects can only be reported in free text format and ontology terms are not natively embedded into IUCLID. However, in this work, the aim was to report effects with appropriate ontology. For this purpose, data were initially extracted by pharmacological reviews and SPLs by using existent ontologies including Ontology Lookup Service (OLS) (OLS, 2022) and Unified Medical Language System (UMLS) (Unified Medical Language System). However, it was evident from the beginning that a merge of ontologies needed to be done and many different terms needed also to be additionally encoded since they were non-existent in other ontologies.

The optimal solution was found by using the human phenotype (HP) and the mammalian phenotype (MP) ontologies of the Ontology Lookup Service (OLS) (OLS, 2022). These were used as a starting point for the development of a hierarchical ontology by using a specific IT tool described below and allowed to create novel and unique standardized terms, including plurals and synonyms.

New ontological terms were then used to populate the effect fields of IUCLID as described above.

2.4.2. Use of ontology terms in IUCLID

The new ontology was built with a hierarchical structure to cover only effects coming from pharmacological reviews and SPLs. The structure has been adopted to allow concepts to be classified into different nodes as depicted in Fig. 2. While some nodes like effect classes (*i.e.* mortality, clinical signs, haematology) or effect systems (*i.e.* gastrointestinal trat, immune system and cardiovascular) are available in IUCLID as a picklist, other nodes needed to be integrated and, in this work, were complemented with additional information relating to effect organs, effect parameters, effect type and other ontology specifications. The newly created ontology can be exported in OWL format and will be made available through the OLS website.

2.5. Data processing

2.5.1. Standard operating procedure

A standard operating procedure (SOP) as depicted in Fig. 3 was developed to reflect all the necessary steps from the pdf data extraction to the creation of a IUCLID dataset. The SOP was created as a practical workflow that includes the retrieval of the documents for a given NDA (typically in pdf format), the selection of relevant studies or sections, their conversion into text files, the annotation of the text files and the transfer of key information into the harmonised templates of IUCLID, complementing them, when necessary, with the tailor-made ontology. The SOP relies on several tools, some of which were developed for the purposes of this project and are described in detail below.

2.5.2. Conversion of the study text into text form

The document blocks containing the information of interest were converted into text with ABBYY fine reader V.15 (ABBYY FineReader PDF, 2022). This step always included optical recognition followed by a manual correction and required fine-tuning for the reliable recognition of special characters as described in the supplementary information (Fig. S1). Figures and tables appearing in pharmacological reviews or SPLs were lost once converted into text and hence were processed manually so that the information is transferred into IUCLID. For completeness and reproducibility, for each endpoint study records a pdf document was attached including the result of the textification and also the original tables and figures as figures; this pdf document allows text extraction (other than from the figures and tables) without further optical recognition.

The attached documents are not amenable to automated data processing but can be informative when individual studies are assessed and increases the transparency of the data structuring operations. Moreover, it allows future improvements of the IUCLID datasets in case of errors or incomplete data transfer into the harmonised templates. In addition, ECHA has developed tools to index and search IUCLID attachments, which means that the hazard assessor is able to also search within the original content, *e.g.* for effects using the original and not the ontology terms (IUCLID Text Analytics, 2022). For completeness, the full original content from US FDA, without any processing, is available upon request.

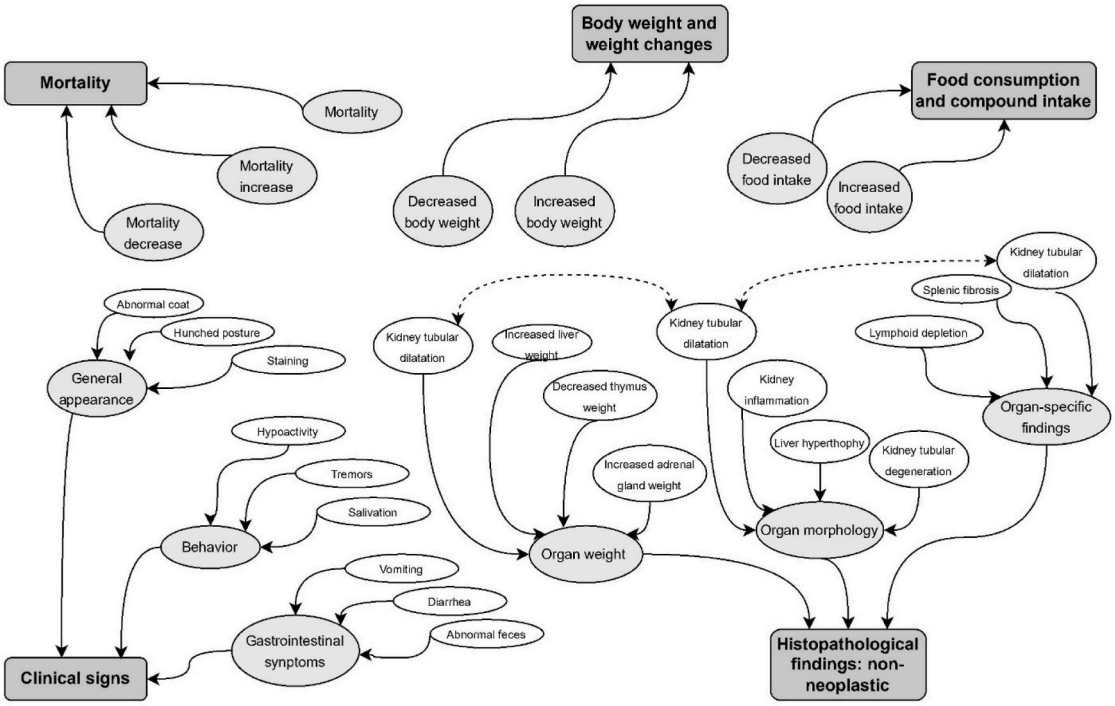

Fig. 2. Graphical representation of the animal and human ontology. Dark grey rounded rectangle represents the 1st level nodes; light grey ellipses 2nd level nodes and white ellipses 3rd level nodes. Dotted line connectors shows an example of effect (i.e. kidney tubular dilatation) belonging to different node paths.

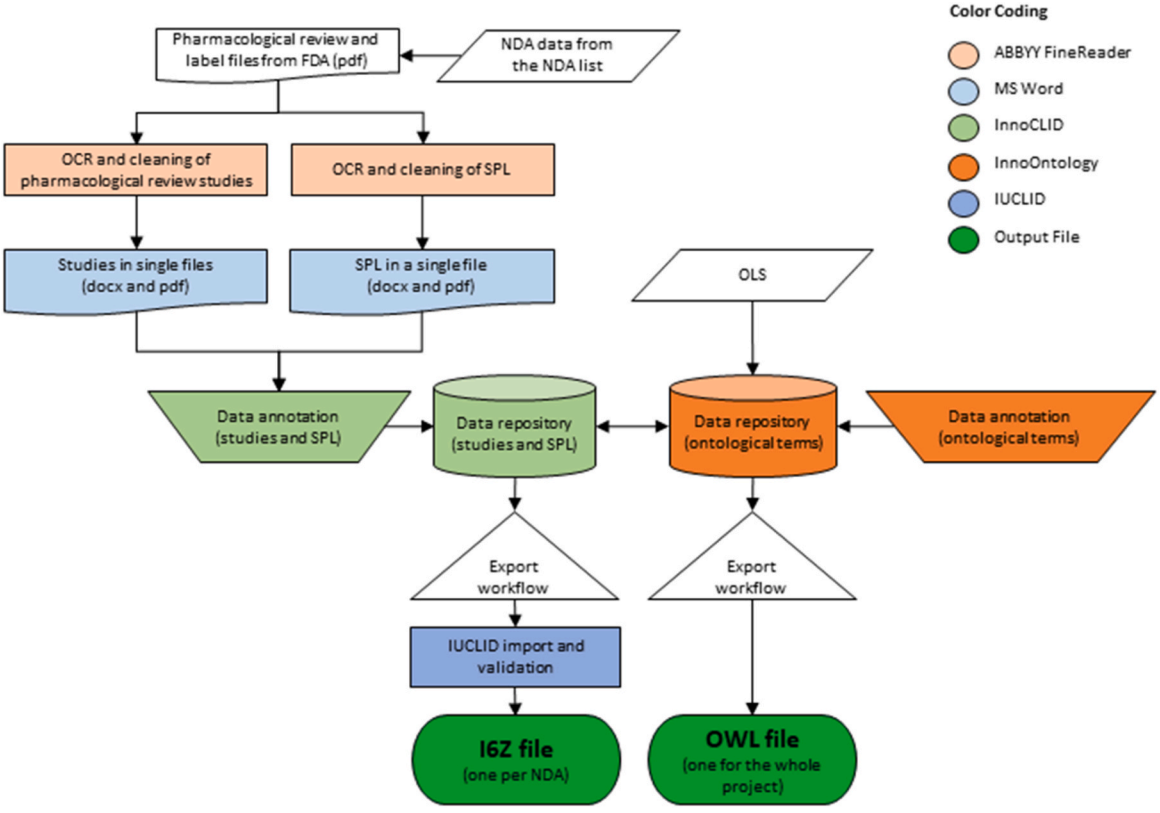

Fig. 3. Standard operating procedure developed for creating IUCLID datasets from pharmacological reviews and SPLs of NDAs.

2.5.3. Data entry

2.5.3.1. Database to capture pharmacological reviews and SPL information. The study information was not entered directly in IUCLID due to the need to standardise terms that are not yet available in the harmonised template picklists. Hence, the study information was first entered into two tailor-made databases through the internally developed user interfaces (UI). The first database (Fig. 4) was created to enter

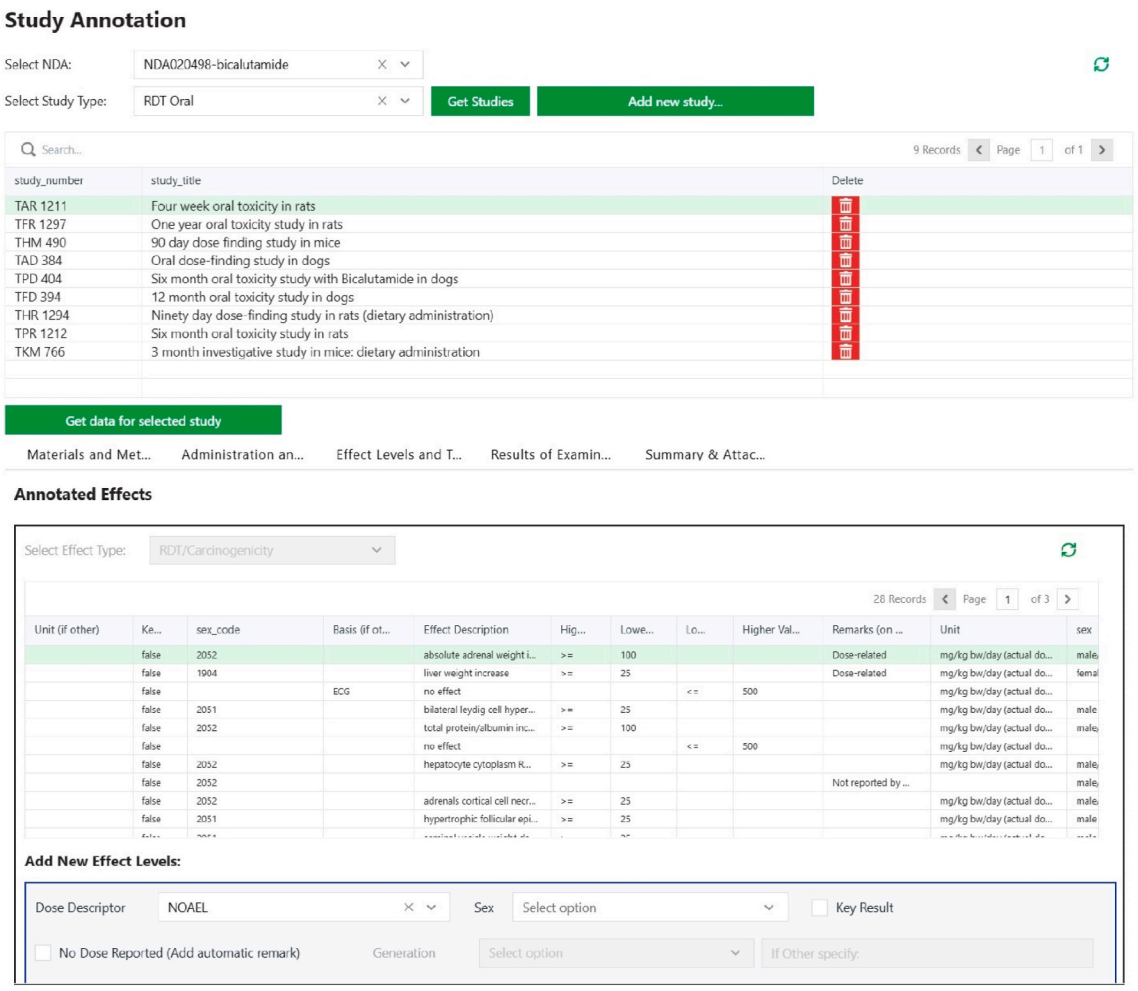

Fig. 4. User interface for data entry of pharmacological and standard product label (SPL) information.

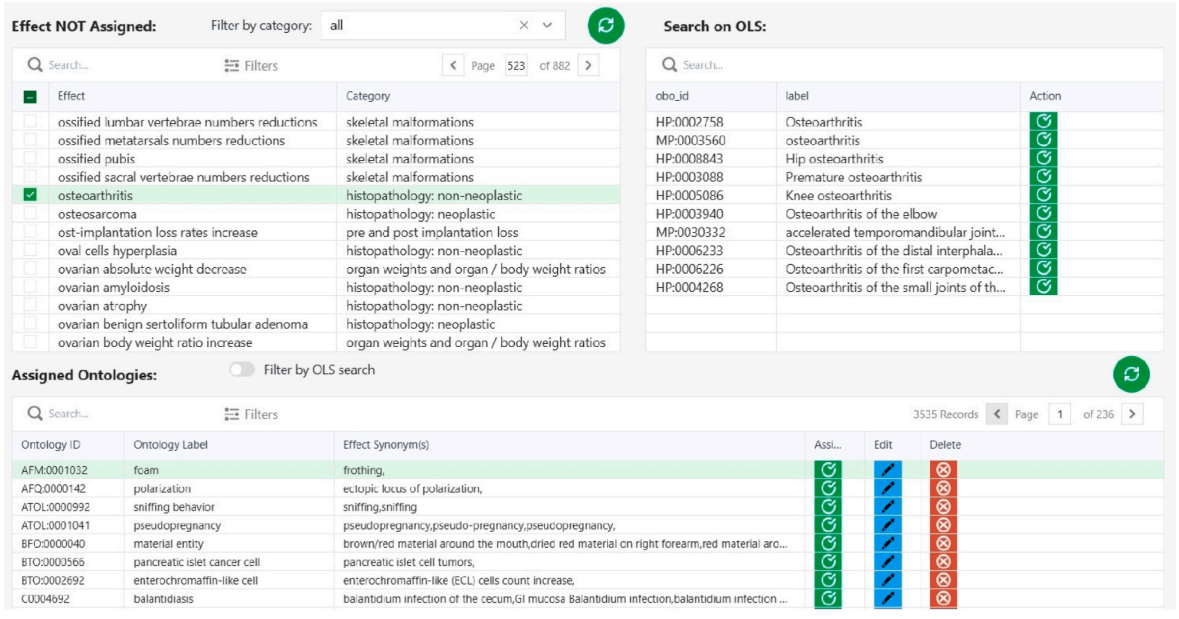

Fig. 5. User interface for data curation of ontological terms.

information retrieved from the pharmacological review and SPL documents. The underlying persistency layer is PostgreSQL (PostgreSQL, 2022) combined with a personalised web-based create, read, update and delete (CRUD) interface for data editing, presentation and management developed with the AppSmith framework (AppSmith). This infrastructure allows data to be stored and collated in a relational local database reflecting the IUCLID data model by using one table per harmonised template as explained in Fig. S2 in the supplementary information. The study and SPL database ensured that reference substances and test material information was created only once and linked to all relevant endpoint study records using document UUIDs as in IUCLID that facilitated the creation of IUCLID datasets as shown in Tables S2–4 of the supplementary information. This intermediate database facilitated the data entry and provided flexibility in fine-tuning the creation of the IUCLID datasets by allowing the use of different filters, aggregations and mappings to IUCLID fields, without the need to re-enter the information in IUCLID. This flexibility was essential because the SOP for creating the IUCLID datasets evolved throughout the project as more NDAs were processed.

2.5.4. Database for storing ontology

The second database was used to store the developed ontology and is shown in Fig. 5. The database was built using the same framework (PostgreSQL/Appsmith) as the first database. The database is integrated with the OLS and allows searching for terms, including the mammalian and the human phenotypes (Ontology Lookup Service, 2022). Whenever a term is encountered that is new or not recognised, the user interface allows custom ontological terms to be created. This process ensured that existing ontologies are reused to the extent possible and only complemented when necessary. For each effect level annotated in the database the term reported in the field “Basis for effect level” was standardized by assigning a suitable ontology term either from OLS or manually from our experts. Details on OHTs fields modified for different types of toxicity studies are reported in supporting information (Table S5). Additional details on ontology creation are presented in the results section.

2.6. Creation of IUCLID datasets

The intermediate database with the study information from the pharmacological reviews and SPLs allows the creation of JSON documents with the study information that are subsequently used for uploading the information into IUCLID (version 6.14.3). Due to a large number of created studies, the extraction of processing of the JSON documents was carried out using the UI interface (see Fig. S3) and tailor-made scripts. The scripts compiled the dataset and uploaded it into IUCLID through the IUCLID API, together with the attachments. Microsoft Office 365 for Enterprise (2022) was used to produce working Excel and Word annotation files.

2.7. Creation of a standalone ontology

The new ontology was developed by using the IT tool depicted in Fig. 5. From this tool, files were imported in Protégé (V.5.5.0) which was used to create the newly developed ontology in OWL file (Musen, 2015). In IUCLID database, this ontology is reflected by using the effect fields of the exact definition reported in OWL file.

2.8. Quality control

Quality control checks were carried out throughout the SOP execution to guarantee data accuracy and consistency. For instance, using Abbyy software (ABBYY FineReader PDF, 2022), low-confidence characters were found and corrected during the optical character recognition stage, with manual corrections as necessary. In addition, the automatic spell checker was used to eliminate any misspelt words after text was collected in Microsoft Word. Furthermore, an additional layer of quality control was provided by the user interfaces as shown in Figs. 4 and 5. These were designed to unlock specific fields only after the previous were filled in. This approach ensured that adverse effects were properly classified, with mandatory fields requiring the classification of each effect. Finally, the use of the ontology for effect classification not only eliminated typing errors but also resolved issues with singular/plural forms as well as synonyms.

3. Results

3.1. Overview of study results

Overall, the IUCLID datasets contained 3357 studies for RDT (2259 oral, 57 dermal, 80 inhalation, 961 other), 516 for carcinogenicity, 1405 for developmental toxicity and 889 for reproductive toxicity, covering 530 substances. Fig. 6 gives an overview of the structured toxicity studies in the form of a heatmap. The heatmap is divided according to the anatomical group of the drug, showing the most substance-rich anatomical groups separately. Each section gives the number of drugs processed and their distribution in terms of study density. As an illustration, looking at the first section, *i.e.* antineoplastic and immunomodulating agents, the collected information related to 128 NDAs. When it comes to oral RDT studies, it is interesting to note that 85 drugs have between 2 and 10 studies, which reflects the importance of these kinds of studies in toxicological evaluation. Equally, it is worth noting the high number of NDAs that have at least one developmental toxicity study, while reproductive studies are less frequent. On the other hand, RDT studies in the dermal and inhalation routes are rare for this anatomical group and in general. The graph also shows a relatively high number of drugs with carcinogenicity information, which further corroborates the added value of the project given the general scarcity of such studies for industrial chemicals. Another interesting piece of information relates to the fact that the representation of studies is dependent on the anatomical class, for instance, in the alimentary tract and metabolism as well as in the nervous system NDAs, it is possible to highlight a higher representation of developmental and reproductive toxicity studies.

Fig. 6 shows the study density, which is not informative when it comes to the number of substances that have toxicity studies of different types. Fig. 7 shows the number of substances with different combinations of study types. For simplicity, the figure does not include a breakdown according to the anatomical group, and pools together all RDT studies regardless of the route of exposure. The Figure shows that 186 substances have a complete dataset, *i.e.* they have at least one of each study type (*i.e.* RDT, carcinogenicity, developmental toxicity and reproductive toxicity study), whilst only a small number of substances contains one study type. This contrasts with the lower number of higher-tier experimental data for industrial chemicals, due to the use of adaptations by registrants, in particular weight-of-evidence and read-across. For instance, the fourth report under Article 117 (3) of the REACH Regulation presents a detailed comparison of the percentage of substances for which registrants have provided at least one guideline study for each information requirement (AppSmith: Open). The results demonstrate a significant use of adaptations to fulfil information requirements, with only 5% of the approximately 2000 fully registered substances at $>$ 1000 tpa having a carcinogenicity study.

3.2. Application of the standard operating procedure

The application of the above-mentioned SOP is exemplified in Fig. 8 where a three-month RDT study for the drug paliperidone was structured and imported into IUCLID. Note that effects and effects levels are entered in a fully structured format.

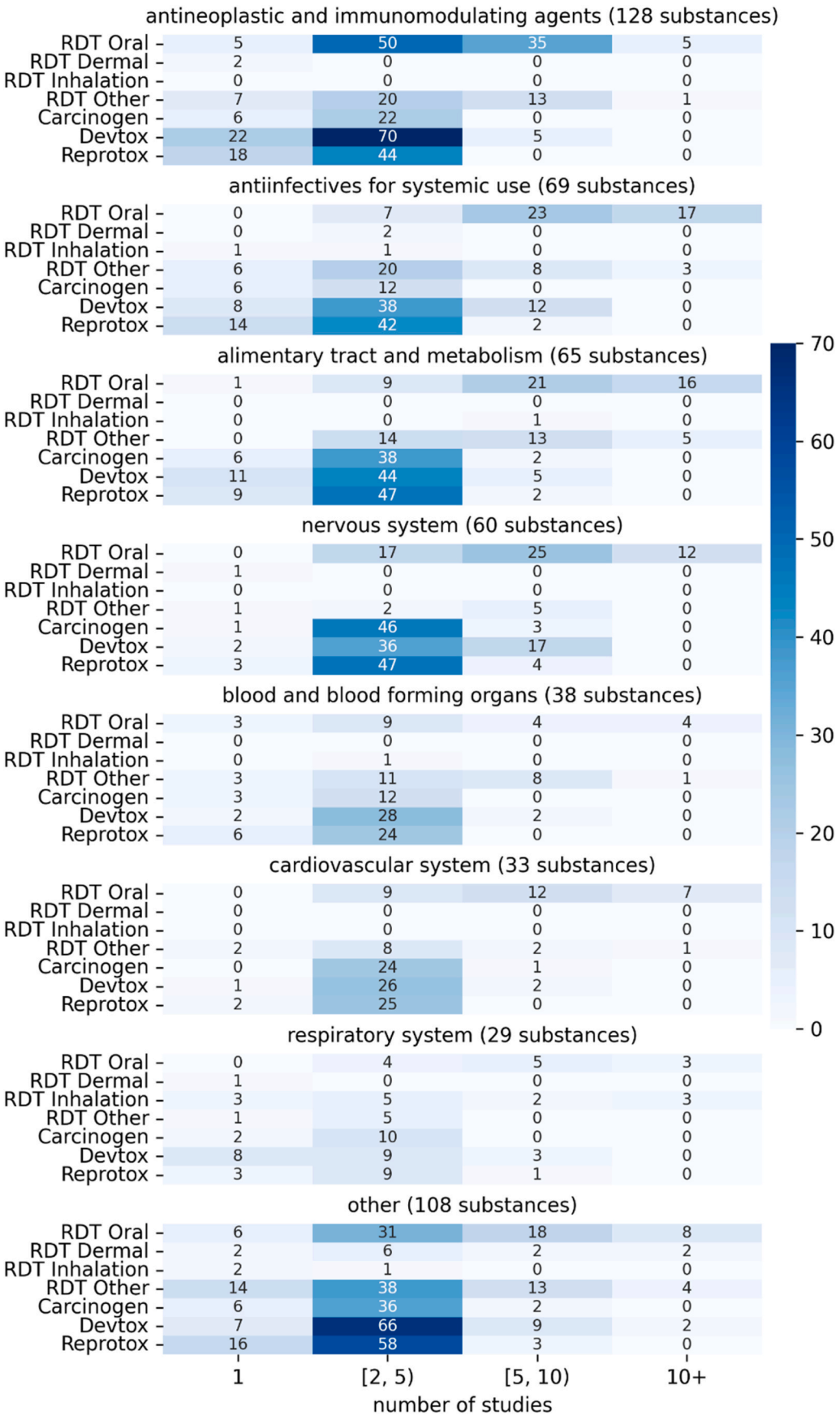

Fig. 6. Study density per ATC anatomical group and toxicity study type.

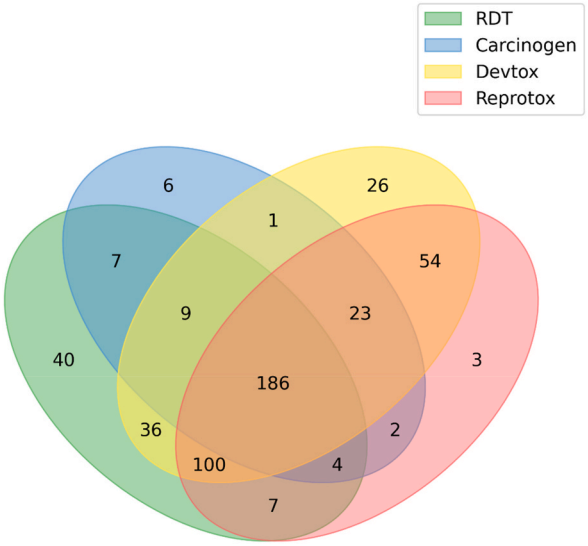

Fig. 7. Venn diagram showing the number of substances for which studies of different types are simultaneously available.

3.3. Uses of the data collected in IUCLID

The collection of data obtained in this work is provided as a set of i6z files at https://iuclid6.echa.europa.eu/us-fda-toxicity-data. A subset of about 350 NDAs is currently released and additional data will be updated within 2023. The i6z files can be imported into IUCLID (desktop or cloud services) by dragging onto the IUCLID user interface either in full or for a selected substance using the accompanying substance index. IUCLID can then be used to browse data, taking advantage of its web interface.

Another possibility is to use the text analytics search engine available from ECHA's website (IUCLID Text Analytics, 2022). Text analytics could be used, for instance, to conduct sophisticated searching of all IUCLID fields including the text content of attachments. In the context of this work, this tool could be useful to search specific effects (*e.g.* late resorptions, follicular thyroid carcinoma) to retrieve entries corresponding to this effect. Interestingly, searches can be carried out on structured data such as picklists and dates, as well as on unstructured data such as free text fields and attachments that are included in the NDA dossiers.

In addition, US FDA data can be extracted in bulk and in a straightforward manner with the IUCLID data extractor. This is an advanced tool that extracts data from IUCLID in accordance with a set of user-defined rules (IUCLID Data Extractor, 2022). It is installed separately from, and then connected to, an instance of IUCLID Server and has its own web-based user interface, separate from that of IUCLID, but modelled on the IUCLID data structure. IUCLID data extractor can be used to export the full array of 530 NDAs for custom data mining and data analysis, using Python, R, Knime (Knime) or another data processing language or tool. IUCLID automatically migrates the data into the latest IUCLID version and the IUCLID corresponding document definitions, and hence the compiled IUCLID datasets can be imported into a future IUCLID instance with no additional conversions needed.

Finally, the data retrieval and aggregation (TEDRA) plugin can be used for data integration between a IUCLID6 server and the QSAR Toolbox (TEDRA Plugin, 2022), ultimately allowing the QSAR Toolbox to read data structured in this work and apply them to hazard assessment. The QSAR Toolbox can also be used for correlating other data in its databases with the structured NDA data and in combination with profiling results and predicted metabolites. These uses will be matter of a future publication.

4. Discussion and conclusions

This work analysed 530 new drug application (NDA) dossiers from the US FDA and integrated animal unstructured data and human information into a new structured database using IUCLID and a tailor-made ontology. This is made possible by the creation of an integrated workflow that started from the digitisation of the original pdf files of pharmacological reviews and standard product labels (SPLs) by means of a standard operating procedure (SOP) to annotate data with an *ad hoc* ontological dictionary.

Although this study has been mainly limited to Type 1 NDAs and specific study types (RDT, carcinogenicity, developmental and reproductive toxicity), it is demonstrated that the workflow can effectively integrate complex data in a structured IUCLID format.

Considerable work has been dedicated to the correct digitisation of data. This step was needed to solve different issues related to the complexity and diversity of sourcing data. Indeed, the original pdf

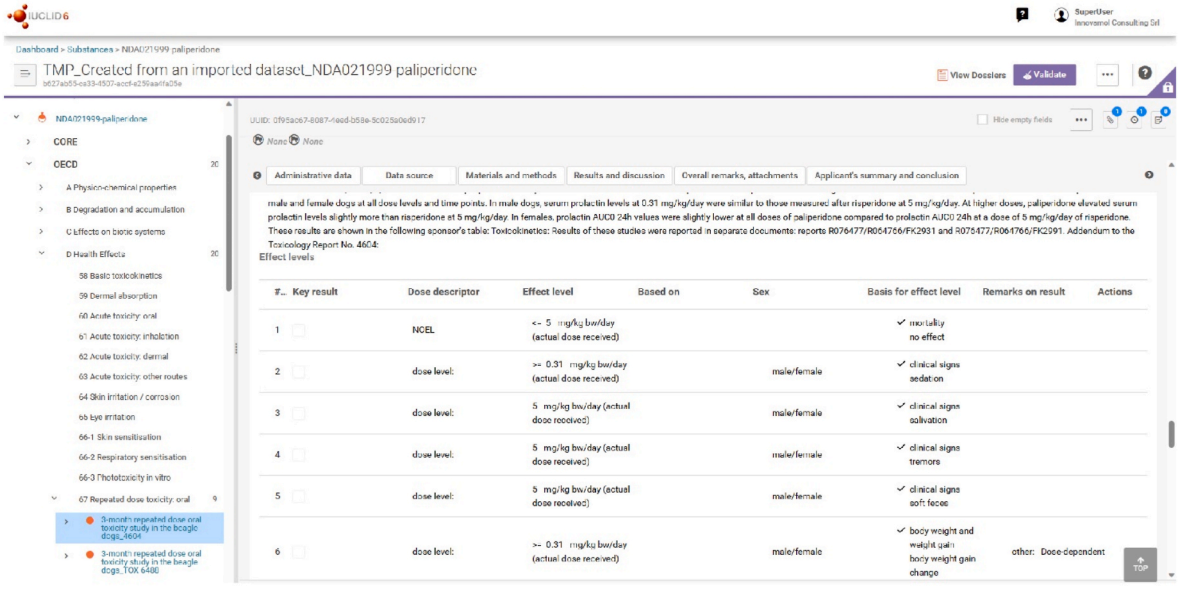

Fig. 8. Example of how effects and effect levels appear in IUCLID for a 3-month RDT study with paliperidone.

spanned from very old pharmacological reviews (*e.g.* from the 1960s) to most recent and easy-to-handle documents. For old documents, the optical character recognition (OCR) capability was mainly focused on the correct recognition of symbols and on the quality improvement of original documents that, in some cases, were also handwritten or typewritten. This work allowed the creation of a custom-trained pattern that can be useful to digitise other old scientific documents, especially those relating to pharmacological or clinical data.

The creation of a rationale and a SOP for linking effects with effect levels was a challenging part of this work. Indeed, the definition of specific rules for annotating numerical data was recognised to be of the utmost importance for ensuring the adequate quality of the final database. For instance, the basis for pharmacological definitions included in the SOP was precisely identified as shown in the result part because medical writers often have different notations or ways to report data. For example, the annotation of the NOAEL, which represents the highest dose level that does not produce an adverse effect compared to the control group (FDA, 2005), required some interpretation because medical writers often referred to NOAEL without reporting it explicitly. In other cases, the distinction between NOAEL and NOEL was not clear, where the latter represents the no-observed effect levels (NOEL), *i.e.* referring to a non-adverse effect. These cases also required some interpretation which was dependent on the wider context of the sentence or the paragraph. When a NOAEL was not indicated by the medical writer, the correspondent annotation was "not reported", thus clearly indicating in the database that the study did not reveal the level of exposure of that drug. A different case, also reported, was when the NOAEL was not indicated because the available data were not sufficient to derive this value, in which case it was reported as "not determined". It is interesting to note that a significant proportion of studies had a not reported or not determined NOAEL.

A precise scheme was devised to assign effect levels with the appropriate mathematical operator (*e.g.* anogenital distance change occurring at dose greater than 3 mg/kg bw/day), since each effect level was associated to a animal or human effect. In several cases, the assessment of the assigned mathematical operator required a small extent of interpretation because the medical writer often described effects and effect levels in text form. Equally, regarding reproductive and developmental studies, it was thoroughly analysed how to classify the studies because the summaries often did not explicitly state the key objectives of testing. It was decided that if the study only evaluated parameters related to sexual function and fertility, or if its main emphasis was on sexual function and fertility, the study was defined as a reproductive toxicity study, for example resembling the OECD test guidelines 443 (Hellsten et al., 2023; OECD, 2018a) and 416 (OECD, 2001). On the other hand, if the main emphasis was on the developmental toxicity of foetuses, the study was defined as a developmental toxicity study, for example resembling the OECD test guidelines 414 (OECD, 2018b). Special care was taken to state the dosing duration clearly. Indeed, there are differences in exposure duration between males and females, also considering that developmental toxicity studies often include only females, *i.e.* dams/does are dosed during gestation, and foetuses are investigated for growth, survival, variations, and malformations.

A major milestone and a challenging part of the present work was the creation of a new ontology for structuring the effects. The goal was to create an ontology embedding not only a dictionary of terms but also their semantic meaning, *i.e.*, capturing the relationships between terms and clarifying their meaning to ensure accurate interpretation. In fact, the developed database contains both animal data and human information, thus using ontologies focusing only on human and/or mammalian phenotypes without integration was insufficient to account all effects. Indeed, it became evident while collecting the first NDAs data that several effects could not be located in the merged OLS human and mammalian phenotype ontology. Therefore, by following the same hierarchical system, additional terms were added to the ontology to allow establishing correlations between observed effects or their absence at different level of granularity. This correlation would not be possible if the effects were described using free text. In addition, correlations may not be observed when using the originally reported terms but could be established when the effects are all expressed at less granular level. Similarly, the use of a hierarchical ontology may facilitate the development of predictive models that go beyond a simplistic toxic/non-toxic outcome that is not sufficient for hazard and risk assessment in a regulatory context. The model fitting may be attempted using effects expressed at different ontology levels, to strike a balance between the detail in which adverse effects can be predicted from structure, *i.e.* more granular ontology terms, and the inherent difficulty to develop predictive models when the training set contains few substances for some of the possible prediction outcomes. As a result, the created ontology might be considered as an improvement of IUCLID, which already provides a solid data structure. In addition, having used publicly available data source and OWL file as output, we aimed to ensure interoperability where different databases can exchange, interpret, and integrate IUCLID data.

This work is not only useful because it increases the amount of toxicity information that exists in a structured and algorithmically processable form, but also because it introduces a comprehensive methodology for structuring legacy toxicity studies that currently exist only in documents. In addition to the methodology and the standard operating procedure, the developed ontology, the workflows and the output files in IUCLID format may be beneficial to legacy toxicity data holders who may wish to use the developed methodology within their own data structuring procedure. Second, because this project focused on substances for which both animal information and human data are simultaneously available, the database can constitute a crucial starting point for the translational exercise to predict human adverse events from animal experimentation. Indeed, it should be recognised that many animal models have been developed through long-standing experience and have been recognised in international regulations and good laboratory practice (GLP). However, despite the availability of a large amount of data coming from decades of animal experimentation, effective data analysis bridging adverse events observed in animal and human experimentation requires the creation of structured datasets as described in this work.

Similarly, the dataset can become a valuable starting point for improved QSAR or quantitative structure-property relationship (QSPR) models considering later stages of pharmaceutical development, for instance, spanning through non-clinical to clinical stages. This is particularly relevant for pharmaceutical companies and contract research organisations (CROs) that have an interest in identifying early adverse reactions preluding to clinical trial failures leading to pharmaceutical development attrition rate. Likewise, this early failure recognition would have important ethical and regulatory implications in terms of reducing test animal use also by means of new approach methodologies (NAMs) based on data sharing, data gathering, and use of computational prediction tools.

Funding

This work was funded by ECHA under the framework contracts ECHA/2011/18 and ECHA/2021/67.

CRediT authorship contribution statement

Martina Evangelisti: Conceptualization, Validation, Formal analysis, Data curation, Writing – original draft. **Marco Daniele Parenti:** Conceptualization, Methodology, Software, Writing – original draft. **Greta Varchi:** Validation, Formal analysis, Data curation. **Jorge Franco:** Validation, Formal analysis, Data curation. **Jochen vom Brocke:** Validation, Formal analysis, Writing – review & editing. **Panagiotis G. Karamertzanis:** Validation, Formal analysis, Visualization,

Writing – review & editing. **Alberto Del Rio:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Ingo Bichlmaier:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Martina Evangelisti and Jorge Franco report an employment relationship with Innovamol Consulting Srl. Alberto Del Rio reports a relationship with Innovamol Consulting Srl that includes: employment, equity or stocks, and funding grants. Jochen vom Brocke, Panagiotis G. Karamertzanis and Ingo Bichlmaier report a relationship with the Europan Chemical Agency that includes: employment and funding grants.

Data availability

We shared the link to our data/code in the manuscript

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yrtph.2023.105416.

References

ABBYY FineReader PDF. https://pdf.abbyy.com/, 2023.

AppSmith: open-source, low-code platform to build, ship, and maintain internal tools [Internet]. https://www.appsmith.com/.

Bodenreider O. Using SNOMED CT in Combination with MedDRA for Reporting Signal Detection and Adverse Drug Reactions Reporting.

Cai, M.C., Xu, Q., Pan, Y.J., Pan, W., Ji, N., Li, Y.B., et al., 2015. ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. Nucleic Acids Res. 43 (D1), D907–D913.

Center for drug evaluation and research. NDA Classification Codes 10, 2015.

Council, 1993. E. U. Council Regulation (EEC) No 793/93 of 23 March 1993 on the Evaluation and Control of the Risks of Existing Substances. COUNCIL REGULATION (EEC) No 793/93. Official Journal of the European Communities.

DrugBank online. https://go.drugbank.com/classyfication, 2023.

Drugs@FDA, 2023. FDA-approved drugs. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm.

Drugs.com [internet]. Drugs.com | prescription drug information, interactions & side effects. https://www.drugs.com/, 2022.

ECHA, 2022a. What is IUCLID? - ECHA [Internet]. https://echa.europa.eu/support/registration/creating-your-registration-dossier/what-is-iuclid-.

ECHA, 2022b. Understanding REACH - ECHA [internet]. https://echa.europa.eu/regulations/reach/understanding-reach.

EMA, 2018. Download Medicine Data [Internet]. European Medicines Agency. https://www.ema.europa.eu/en/medicines/download-medicine-data.

European Parliament, 2008. REGULATION (EC) No 1272/2008 of the EUROPEAN PARLIAMENT and of the COUNCIL of 16 December 2008 on Classification, Labelling and Packaging of Substances and Mixtures, Amending and Repealing Directives 67/548/EEC and 1999/45/EC, and Amending Regulation (EC) No 1907/2006. fficial Journal of the European Union.

European Parliament and Council, 2012. Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 Concerning the Making Available on the Market and Use of Biocidal Products, vol. 167, pp. 1–123.

European Parliament, Council of the European Union, 2006. Regulation (EC) No 1907/2006 of the European parliament and of the Council of 18 december 2006 concerning the registration, evaluation, authorisation and restriction of chemicals (REACH), establishing a European chemicals agency, amending directive 1999/45/EC and repealing Council regulation (EEC) No 793/93 and commission regulation (EC) No 1488/94 as well as Council directive 76/769/EEC and commission directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC [internet]. 32006R1907 2006. http://data.europa.eu/eli/reg/2006/1907/oj.

FDA, 2005. Estimating the maximum safe starting dose in initial clinical trials for therapeutics in adult healthy volunteers. Pharmacol. Toxicol. 30.

FDA, 2019. FDA Established Pharmacologic Class.

FDA, 2023. Drugs@FDA: FDA-approved drugs [internet]. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm.

Heidorn, C.J.A., Rasmussen, K., Hansen, B.G., Nørager, O., Allanou, R., Seynaeve, R., et al., 2003. IUCLID: an information management tool for existing chemicals and biocides. J. Chem. Inf. Comput. Sci. 43 (3), 779–786.

Hellsten, K., Suchanová, B.B., Sihvola, V., Simanainen, U., Leppäranta, O., Chronis, K., et al., 2023. The importance of study design in investigating intrinsic developmental toxic properties of substances in new studies under the REACH and CLP Regulations in the European Union. Curr Opin Toxicol, 100402.

Holmgren, S.D., Boyles, R.R., Cronk, R.D., Duncan, C.G., Kwok, R.K., Lunn, R.M., et al., 2021. Catalyzing knowledge-driven discovery in environmental health sciences through a community-driven harmonized language. Int. J. Environ. Res. Publ. Health 18 (17), 8985.

IUCLID data extractor [internet]. https://iuclid6.echa.europa.eu/en/data-extractor, 2023.

IUCLID text analytics [internet]. https://iuclid6.echa.europa.eu/en/text-analytics, 2023.

Ives, C., Campia, I., Wang, R.L., Wittwehr, C., Edwards, S., 2017. Creating a structured adverse outcome pathway knowledgebase via ontology-based annotations. Appl Vitro Toxicol 3 (4), 298–311.

Knight, D.J., Deluyker, H., Chaudhry, Q., Vidal, J.M., de Boer, A., 2021. A call for action on the development and implementation of new methodologies for safety assessment of chemical-based products in the EU – a short communication. Regul. Toxicol. Pharmacol. 119, 104837.

Knime [internet]. https://www.knime.com/.

Knudsen, T., Martin, M., Chandler, K., Kleinstreuer, N., Judson, R., Sipes, N., 2013. Predictive models and computational toxicology. In: Barrow, P.C. (Ed.), Teratogenicity Testing [Internet], Methods in Molecular Biology, vol. 947. Humana Press, Totowa, NJ, pp. 343–374 [cited 2023 Apr 28]. https://link.springer.com/10.1007/978-1-62703-131-8_26.

Musen, M.A., 2015. The Protégé project: a look back and a look forward. AI Matters 1 (4), 4–12.

OECD, 2001. Test No. 416: Two-Generation Reproduction Toxicity [Internet]. OECD (OECD Guidelines for the Testing of Chemicals, Section 4). Available from: https://www.oecd-ilibrary.org/environment/test-no-416-two-generation-reproduction-toxicity_9789264070868-en (OECD Guidelines for the Testing of Chemicals, Section 4). Available from:

OECD, 2018a. Test No. 443: Extended One-Generation Reproductive Toxicity Study [Internet]. OECD (OECD Guidelines for the Testing of Chemicals, Section 4). https://www.oecd-ilibrary.org/environment/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264185371-en.

OECD, 2018b. Test No. 414: Prenatal Developmental Toxicity Study [Internet]. OECD (OECD Guidelines for the Testing of Chemicals, Section 4). Available from: https://www.oecd-ilibrary.org/environment/test-no-414-prenatal-development-toxicity-study_9789264070820-en.

OLS, 2022. Ontology Lookup service < EMBL-EBI [internet]. https://www.ebi.ac.uk/ols/index.

Ontology Lookup service < EMBL-EBI [internet]. https://www.ebi.ac.uk/ols/index.

PostgreSQL [internet]. https://www.postgresql.org/, 2022.

QSAR Toolbox [internet]. https://qsartoolbox.org/support/ontologies, 2023.

TEDRA plugin [internet]. https://qsartoolbox.org/download/#tedra, 2022.

Unified Medical Language System (UMLS) [internet]. https://www.nlm.nih.gov/research/umls/index.html.

Watford, S., Edwards, S., Angrish, M., Judson, R.S., Paul Friedman, K., 2019. Progress in data interoperability to support computational toxicology and chemical safety evaluation. Toxicol. Appl. Pharmacol. 380, 114707.