Deliverable D3.4

Final database with additional content

Grant Agreement	Health-F5-2008-200787
Acronym	OpenTox
Name	An Open Source Predictive Toxicology Framework
Coordinator	Douglas Connect

Contract No.	Health-F5-2008-200787	
Document Type:	Deliverable Report	
WP/Task:	WP3/ D3.4	
Name	Final database with additional content	
Document ID:	OpenTox Deliverable Report WP3-D3.4	
Date:	August 31, 2011	
Status:	Revised draft	
Organisation:	Istituto Superiore di Sanità, Rome, Italy	
Contributors	Romualdo Benigni	ISS
	Olga Tcheremenskaia	ISS
	Vedrin Jeliazkov	IDEA
	Nina Jeliazkova	IDEA
	Ivelina Nikolova	IDEA
	Sylvia E. Escher	ITEM
	Monika Batke	ITEM
	Tomas Baier	ITEM
	Barry Hardy	DC
	Roman Affentranger	DC

Distribution:	Public

Purpose of Document:	To document results for this deliverable

Document History:	1 -Initial draft distributed on May 02, 2011
	2 – Revised draft distributed on May 26, 2011
	3 – Revised draft distributed on May 30, 2011
	4 – Revised draft distributed on August 18, 2011
	5 – Revised draft distributed on August 21, 2011
	6 – Revised draft August 31, 2011

Table of Contents

Table of Figures 6
List of Tables 7
Summary 8
1. OpenTox database 10
1.1 Overview 10
1.2 Linked resources 11
1.3 Implementation independent architecture 11
1.4 AMBIT database and software 12
1.4.1 Chemical compounds 12
1.4.2 Data provenance and quality assurance 13
1.4.3 Identifiers, descriptors and properties 14
2. Datasets included in the final database 15
2.1 Selection of data sources for inclusion 15
2.2 Included datasets details 21
2.2.1 ECHA list of pre-registered substances 22
2.2.2 Chemical Identifier Resolver 22
2.2.3 ChemIDplus 22
2.2.4 ChemDraw 22
2.2.5 JRC PRS list 23
2.2.6 ISSCAN 23
2.2.7 ISSMIC 23
2.2.8 ISSSTY 23
2.2.9 CPDBAS 24
2.2.10 DBPCAN 24
2.2.11 EPAFHM 24
2.2.12 FDAMDD 24
2.2.13 HPVCSI 25
2.2.14 HPVISD 25
2.2.15 IRISTR 25
2.2.16 KIERBL 26
2.2.17 NCTRER 26
2.2.18 NTPBSI 26
2.2.19 NTPHTS 26
2.2.20 TOXCST 27

2.2.21 TOXCST_ACEA 27
2.2.22 TOXCST_Attagene 27
2.2.23 TOXCST_BioSeek 27
2.2.24 TOXCST_Cellumen 28
2.2.25 TOXCST_CellzDirect 28
2.2.26 TOXCST_Gentronix 28
2.2.27 TOXCST_NCGC 28
2.2.28 TOXCST_Novascreen 29
2.2.29 TOXCST_Solidus 29
2.2.30 TOXCST_ToxRefDB 29
2.2.31 TXCST2 30
2.2.32 ECETOC skin irritation 30
2.2.33 Skin sensitisation (LLNA) 30
2.2.34 Skin sensitisation (LLNA) – second compilation 30
2.2.35 Bioconcentration factor (BCF) gold standard database 31
2.2.36 Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way 31
2.2.37 Benchmark Data Set for In Silico Prediction of Ames Mutagenicity 31
2.2.38 Bursi AMES Toxicity Dataset 31
2.2.39 EPI_AOP 32
2.2.40 EPI_BCF 32
2.2.41 EPI_BioHC 32
2.2.42 EPI_Biowin 33
2.2.43 EPI_Boil_Pt 33
2.2.44 EPI_Henry 33
2.2.45 EPI_KM 33
2.2.46 EPI_KOA 34
2.2.47 EPI_Kowwin 34
2.2.48 EPI_Melt_Pt 34
2.2.49 EPI_PCKOC 34
2.2.50 EPI_VP 35
2.2.51 EPI_WaterFrag 35
2.2.52 EPI_Wskowwin 35
2.2.53 Chemical Name to Structure Converter (OPSIN) 35
2.2.54 PubChem Structures + Assays 36
2.2.55 Leadscope_carc_level_2 36
2.2.56 Leadscope_ccris_genetox 36

2.2.57 Leadscope_cder_chronic 37
2.2.58 Leadscope_cder_genetox 37
2.2.59 Leadscope_cder_repro_dev 37
2.2.60 Leadscope_cfsan_acute 37
2.2.61 Leadscope_cfsan_chronic 38
2.2.62 Leadscope_cfsan_genetox 38
2.2.63 Leadscope_cfsan_repro_dev 38
2.2.64 Leadscope_fda_marketed_drugs 38
2.2.65 Leadscope_genetox_level_2 39
2.2.66 Leadscope_ntp_genetox 39
2.2.67 Pharmatrope_AERS_hepatobiliary_system 39
2.3 Quality Labels Overview 39
2.4 Fraunhofer ITEM Repdose database 47
3. OpenTox Toxicological Endpoints Ontology 47
3.1 ECHA endpoints ontology 47
3.2 OpenTox Toxicological Ontology 50
3.3 OpenTox Organs and Effects Ontology 51
3.4 Automatic conversion of ToxML schema to an OWL ontology 53
4. Conclusions 54
Appendix A: OpenTox final database raw statistics 56
Appendix B: OpenTox final database structure 71
Appendix C: OpenTox Toxicological Endpoints Ontology 94

Table of Figures

Figure 1 Final database entity-relationship diagram 13
Figure 2 ECHA endpoints ontology summary graph 48
Figure 3 Human health specific part of the ECHA endpoints ontology 49
Figure 4 OpenTox toxicological ontology structure 51
Figure 5 Example of the target site description (Organ System Ontology) for bronchi 52
Figure 6 Overview of the structure of the combined organ (in orange) and effect (in green) ontologies..... 53
Figure 7 Introducing object type properties for each tuple of nested classes..... 54

List of Tables

Table 1 Quality labels mapping *14*
Table 2 Final OpenTox database automatic classification results distribution *40*
Table 3 Final OpenTox database quality labels distribution *45*

Summary

A major prerequisite for the successful implementation of the main principles of the Three Rs Declaration of Bologna, adopted by the 3rd World Congress on Alternatives and Animal Use in the Life Sciences (Bologna, Italy, August 31st 1999) – namely Reduction, Refinement and Replacement alternatives – is the universal access to high-quality experimental data on various toxicological, biological and chemical properties. The range of reduction and replacement alternative testing methods includes the following OpenTox-relevant approaches:

- the improved storage, exchange and use of information from animal experiments already carried out, so that unnecessary repetition can be avoided;
- the use of physical and chemical techniques, and of predictions based on the physical and chemical properties of molecules;
- the use of mathematical and computer modelling, including modelling of structure-activity relationships, molecular modelling and the use of computer graphics, and modelling of biochemical, pharmacological, physiological, toxicological and behavioural processes.

Since in many circumstances an animal test cannot be currently replaced by a single alternative method, the development, evaluation and optimisation of stepwise testing strategies and integrated testing schemes should be encouraged. The OpenTox data facilities, made publicly accessible through a web services framework, provide a solid basis for addressing the above-mentioned reduction and replacement alternatives goals in a more efficient, technically sound and integrated way compared to current uncoordinated practices and fragmented resources. Unfortunately, even today, more than 10 years after the adoption of the Three Rs Declaration of Bologna, the "state-of-the-art" is characterised by highly fragmented and unconnected life sciences data (both from a physical and ontological perspective), which is furthermore frequently inaccurate and/or difficult if not impossible to find or access. The OpenTox approach to data resource management and integration has the following major features, which address the replacement alternatives challenge and associated user, industry and regulatory needs including REACH:

- a universal database structure design, allowing for storage of multi-faceted life sciences data;
- an ontology allowing for efficient mapping of similar and/or complementary data coming from different datasets into a unifying structure having a shared terminology and meaning;
- the integration of multiple datasets with proven high-quality physico-chemical and/or experimental toxicity data;
- built-in heuristics for automatic discovery of 2D chemical structure inconsistencies;
- extensive support for structure-, substructure- and similarity-based searching of chemical structures;
- an OpenTox standards-compliant dataset interface that allows query submission and results retrieval from any OpenTox standards-compliant web service;
- transparent access to and use of life sciences data, hosted at various physical locations and incorporating a variety of distributed software resources, linked through the OpenTox Framework.

The OpenTox final database includes ECHA's list of pre-registered substances[1] along with high-quality data from consortium members (e.g. ISSCAN,ISSMIC and ISSSTY from partner ISS[2], AMBIT from partner IDEA[3]) and

[1] apps.echa.europa.eu/preregistered/pre-registered-sub.aspx accessed on May 27, 2011

[2] www.iss.it/meca/dati/cont.php?id=199&lang=1&tipo=25 accessed on May 27, 2011

[3] ambit.sourceforge.net accessed on May 27, 2011

third parties (e.g. JRC PRS list[4], EPA DSSTox[5], ECETOC skin irritation[6], LLNA skin sensitisation[7,8], Bioconcentration factor Gold Standard Database[9], EPA ToxCast[10], Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way[11], Benchmark Data Set for In Silico Prediction of Ames Mutagenicity[12], Bursi AMES Toxicity Dataset[13], EpiSuite data[14], PubChem[15], Leadscope data[16] and Pharmatrope data[17]). Additional data for chemical structures has been collected from various public sources (e.g. Chemical Identifier Resolver[18], ChemIDplus[19]), generated by chemical name to structure conversion (Opsin[20]), and further checked manually by experts. The database provides means to identify the origin of the data, i.e., the specific inventory or software a compound originated from. The data is currently publicly available and accessible through standardized web services[21], as defined in the OpenTox framework design (except for Leadscope and Pharmatrope datasets, which are accessible exclusively through Authorisation Policies, i.e., restricted to OpenTox partners). We also developed an OWL (Web Ontology Language)[22] ontology of toxicological endpoints, which corresponds to the endpoint classification of REACH guidance documents[23] and allows for a unique mapping between endpoints from various inventories.

This report describes in detail the above-mentioned OpenTox data facilities and resources. Section 1 introduces the main concepts of the OpenTox database and its implementation as a web service. The OpenTox

---

[4] ecb.jrc.ec.europa.eu/documents/QSAR/INFORMATION_SOURCES/EC_CHEMICAL_INVENTORIES/ accessed on May 27, 2011

[5] www.epa.gov/ncct/dsstox/ accessed on May 27, 2011

[6] ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) accessed on May 27, 2011

[7] Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz G, Basketter DA. (2005). Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives. Dermatitis 16(4): 157-202. accessed on May 27, 2011

[8] Kern PS, Gerberick GF, Ryan CA, Kimber I, Aptula A, Basketter DA. (2010). Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. Dermatitis 21(1): 8-32. accessed on May 27, 2011

[9] ambit.sourceforge.net/euras accessed on May 27, 2011

[10] EPA ToxCast: Screening chemicals to predict toxicity faster and better accessed on May 27, 2011

[11] Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way accessed on May 27, 2011

[12] Benchmark Data Set for In Silico Prediction of Ames Mutagenicity accessed on May 27, 2011

[13] Bursi AMES Toxicity Dataset accessed on May 27, 2011

[14] EpiSuite data accessed on May 27, 2011

[15] Downloadable Structure Files of PubChem Compounds accessed on May 27, 2011

[16] http://www.leadscope.com/ accessed on May 27, 2011

[17] http://www.pharmatrope.com/ accessed on May 27, 2011

[18] cactus.nci.nih.gov/chemical/structure accessed on May 27, 2011

[19] chem.sis.nlm.nih.gov/chemidplus accessed on May 27, 2011

[20] Daniel M. Lowe, Peter T. Corbett, Peter Murray-Rust, and Robert C. Glen. Chemical Name to Structure: OPSIN, an Open Source Solution. Chem. Inf. Model., 2011, 51 (3), pp 739-753 accessed on May 27, 2011

[21] http://apps.ideaconsult.net:8080/ambit2/ accessed on May 27, 2011

[22] www.w3.org/TR/owl-features accessed on May 27, 2011

[23] guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?%20vers=20_08_08 accessed on May 27, 2011

database relies on OpenTox partner IDEA's implementation of the OpenTox API – the AMBIT database and software – and its organisation is described in section 1.4. We explain its underlying principles, illustrate how chemical compounds and their properties are handled, how data provenance is enabled and maintained, and describe the quality assurance which results in quality labels for the chemical structures. The datasets included in the final version of the OpenTox database are presented in section 2. Details are given for each of the 67 final datasets originating from various sources described above, as well as for the sources themselves. We also present the results of the automatic quality label classification performed on the imported structures and discuss our findings. In section 3 we discuss the OpenTox Toxicological Endpoints Ontology encompassing the ECHA Endpoints Ontology, the OpenTox Toxicological Ontology, the OpenTox Organs and Effects Ontology, and an ontology based on an automatic conversion of the ToxML schema into an OWL ontology. The final section of this report presents our conclusions drawn from the development and the implementation of the OpenTox final database.

1. OpenTox database

1.1 Overview

The OpenTox database is implemented as an OpenTox dataset web service, and accessible via the OpenTox Application Programming Interface (API)[24]. The OpenTox compound, dataset and feature APIs provide generic means to access chemical compounds and aggregate various data. Chemical compounds are assigned unique URIs, and can be retrieved, created, or deleted via HTTP commands submitted to the OpenTox compound service. The HTTP GET command returns a representation of the chemical compound in a specified MIME format. Changing the MIME format (an internet media type, named after the term "Multipurpose Internet Mail Extension"[25]) returns the representation of the compound in that format, making the service essentially work as a format converter. The supported MIME types for datasets (besides the mandatory *application/rdf+xml*) currently are: *chemical/x-mdl-sdfile*, *text/n3*, *application/x-turtle*, *chemical/x-mdl-molfile*, *chemical/x-cml*, *chemical/x-daylight-smiles*, *chemical/x-inchi*, *text/x-arff*, *application/pdf*, *text/uri-list, text/csv*, *text/plain*. The REpresentational State Transfer (REST) architecture[26] natively supports multiple representations per resource, which allows using the most appropriate format for carrying out a particular task.

The services, implementing the OpenTox API for compounds, datasets and features, enable importing arbitrary files with chemical structures and their properties, allowing linking to computer-readable information about the data fields, as well as keeping provenance information. In addition, they support multiple structures of the same compound, which is useful for storing and working with multiple conformations, as well as for comparing structures originally residing in different source databases. Uploading a file with chemical structures and data makes it automatically available online as an OpenTox dataset in several formats. The concept of a dataset of chemical compounds is central to the OpenTox web services functionality. The datasets are directly used by the entire set of remote OpenTox services, implementing different calculations and applying predictive models. Algorithm services accept a dataset URI (Unique Resource Identifier)[27] to build a model or to generate descriptor values. Model services accept a dataset URI to apply a model and obtain predictions. Predictions are also returned as a dataset URI, whose contents could be subsequently retrieved. Search results (by identifiers,

[24] http://www.opentox.org/dev/apis/api-1.1 accessed on May 27, 2011

[25] http://en.wikipedia.org/wiki/Internet_media_type accessed on October 27, 2011

[26] R. T. Fielding, Representational State Transfer (REST) http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm, accessed on Apr 28, 2011

[27] http://en.wikipedia.org/wiki/URI accessed on October 27, 2011

similarity, or substructure) are available as datasets as well. The OpenTox Dataset (*ot:Dataset class*) can be thought of as a file of chemical compounds along with their properties, which is identified (and referred to) by a unique web address instead of a filename and can be read and written remotely. The dataset HTTP POST operation allows uploading datasets in RDF (W3C Resource Description Framework)[28] representation, as well as files with chemical structures with an arbitrary set of fields. The services do not restrict entering and uploading data to predefined fields only. Instead, arbitrary data can be imported and later annotated to establish the semantics of the fields. The fields (e.g. various properties in *sdf* files or any other file format) can contain arbitrary attributes, which are represented as instances of the *ot:Feature* class from the OpenTox ontology. Every feature is identified by a unique URI, and its RDF representation includes a feature name, units, and a link to the resource that was used to generate this property or where it was originally read from.

1.2 Linked resources

Uploading data into the OpenTox dataset services and running calculations through the OpenTox API generates a multitude of linked resources, all available via their RDF representations. The links can span many remote sites running various implementations of OpenTox services. For example, a model, built by model services running at site A, will be accessible via its web address, but the representation could include links to the training dataset and prediction variables hosted by OpenTox services running at site B. The features, representing predicted variables, contain links back to the remote model and thus enable easy retrieval of the RDF representations from an arbitrary set of geographically distributed OpenTox services. It is equally easy to create a snapshot of the content of a given subset of services of particular interest, either for archiving purposes, or to import it into a RDF triple storage and expose it via a SPARQL endpoint. The services provide dynamically generated RDF representation of chemical compounds and their properties and can be crawled in a similar way as search engines crawl the web. However, there is the additional benefit of being able to retrieve the results and perform reasoning over structured data, instead of just analysing keywords and links the way popular search engines typically operate today.

1.3 Implementation independent architecture

Exposing functionalities through a web application programming interface allows hiding the implementation details of both data storage (different database types vs. memory vs. file system backend) and processing (descriptor calculation and machine learning algorithms, using open source, commercial or in-house implementations). The availability of data and processing resources as RDF facilitates integrating the resources as Linked Data[29]. Illustrations of this concept are the 6 different implementations of the OpenTox algorithm and model APIs developed by OpenTox partners (IST, IDEA, NTUA, TUM, SL, IBMC), as well as the two different implementations of the compound and dataset APIs (IDEA, IST). The IST implementation is mainly used to provide temporary storage for calculation and validation procedures. The OpenTox database provides temporary and persistent storage for toxicity data and relies on IDEA's implementation of the OpenTox API (AMBIT database and software)[30].

[28]Resource Description Framework http://www.w3.org/RDF/, accessed on Apr 28, 2011

[29]en.wikipedia.org/wiki/Linked_Data accessed on May 27, 2011

[30] Nina Jeliazkova, Vedrin Jeliazkov. AMBIT RESTful web services: an implementation of the OpenTox application programming interface. *Journal of Cheminformatics* 2011, **3:**18

1.4 AMBIT database and software

The AMBIT database is a relational database consisting of several repositories for compounds, properties, QSAR models, users, references, as well as several tables containing pre-processed information, which allows speeding up substructure and similarity queries. The current implementation is based on MySQL[31].An overview of the entity-relationship diagram of the database is provided in Figure 1. The database has been optimized and partially denormalized to improve the dataset service performance. Additional fields were added to accommodate support for assigning dataset licenses and rights holders.

1.4.1 Chemical compounds

The chemical compounds are stored in the table *chemicals* and assigned a unique number. If connectivity is available, a unique SMILES[32], as well as InChI[33] and molecular formula are generated and stored. The database supports multiple 3D structures per compound, either coming from different inventories, or generated by external programs and imported into the database. The chemical structures are stored in the table *structure* as a compressed text, where supported formats are SDF[34], MOL and CML[35]. The choice of text format makes the database transparent and easy to be used by external software. Support of multiple formats is motivated by the need to keep the data in the original format. If the original format is none of the above formats, it is converted to MOL.

[31] www.mysql.com accessed on May 27, 2011

[32] www.daylight.com/smiles accessed on May 27, 2011

[33] www.iupac.org/inchi accessed on May 27, 2011

[34] www.symyx.com/downloads/public/ctfile/ctfile.jspaccessed on May 27, 2011

[35] en.wikipedia.org/wiki/Chemical_Markup_Language accessed on May 27, 2011

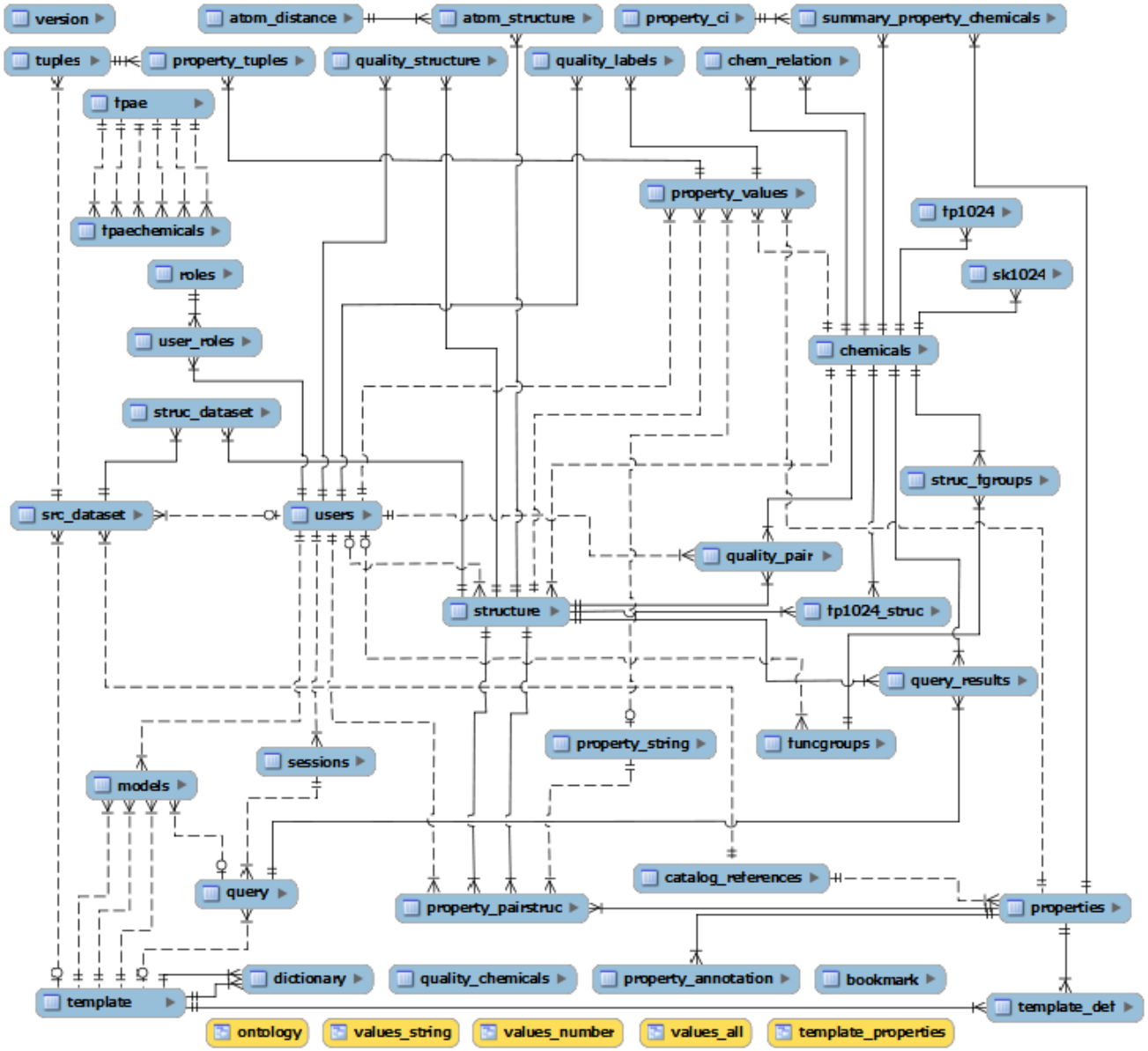

Figure 1 Final database entity-relationship diagram

1.4.2 Data provenance and quality assurance

The database provides means to identify the origin of the data, i.e., the specific inventory a compound originated from. An inventory is identified by its name and reference (table *src_dataset*). Each compound might belong to multiple inventories (table *struc_dataset*), thus allowing users to select the compounds of interest for specific regulatory purposes. Moreover, the data provenance indicator can distinguish between different conformations, for example in cases where a particular conformation of a compound comes from one inventory and a different conformation comes from another inventory.

When structures are imported from a file, they are stored in their original format into the *structure* table. If a structure is subsequently updated as a result of a specific calculation (e.g. 3D conversion), a new version will be added and a pointer to the generation procedure recorded.

The correctness of the chemical structure is crucial for QSAR modelling. In order to raise the awareness of possibly incorrect structures that might be imported from external sources, AMBIT allows assigning quality labels to each 2D chemical structure, as follows:

- Manual verification by expert(s). Any user can assign quality labels and explain the reason of the assignment (table *quality_structure*). The reasons can include discrepancies between registry numbers, names and structure, expert knowledge, manual comparison with external sources, etc. The list of quality labels includes:
    - '*OK*' – the structure is correct;
    - '*ProbablyOK*' – most probably the structure is correct, but some issues still need to be verified;
    - '*Unknown*' – not possible to assign a definite label;
    - '*ProbablyERROR*' – most probably there is an error;
    - '*ERROR*' – the structure is definitely wrong.
- Automatic verification by comparing the structures available under the same chemical compound entry (e.g. imported from different sources) – table *quality_chemicals*:
    - '*Consensus*' – all structures under the same chemical compound entry are identical;
    - '*Majority*' – the majority of structures under the same chemical compound entry are the same, but there are a small number of structures that differ from the majority (e.g. structures come from 3 different sources and two of the three structures are identical);
    - '*Ambiguous*' – there is no majority of equal structures under the same chemical compound entry (e.g. structures come from 3 different sources and all of them are different);
    - '*Unconfirmed*' – the structure comes from a single source and it is impossible to make an automatic comparison.

Furthermore, the results of the automatic comparisons of structures coming from different sources are used for assigning initial quality labels, subject to further review by experts, according to the mapping shown in Table 1.

Table 1 Quality labels mapping

Automatic Classification	Initial Quality Label Assigned
Consensus	OK
Majority	ProbablyOK for the structure that belongs to the majority ProbablyERROR for the structure(s) that belong(s) to the minority
Ambiguous	Unknown (multiple sources)
Unconfirmed	Unknown (single source)

1.4.3 Identifiers, descriptors and properties

The database schema is designed to provide unified storage for an arbitrary number of text strings (e.g. registry numbers or names) and numerical properties (e.g. descriptors, experimental data). The properties are not predefined, but stored in the database on demand, e.g. the AMBIT database is ready to incorporate any number of chemical compounds, identifiers, descriptors and experimental data.

A property (table *properties*) is identified by a name and a reference, thus allowing to distinguish properties with coinciding names, but originating from different sources (e.g. LogP calculated internally by different methods and LogP imported from an external file). Each new property or descriptor is added to the *properties* table with information about the property/descriptor name, units, alias and reference. The reference for a

property imported from a file is the name of the file itself, while the reference for a descriptor contains the name of the software used for calculating it. The alias currently contains a copy of the name, except in cases when the property is recognised as a specific type of registry number or a chemical name. In this case, the alias is assigned a fixed value (e.g. CasRN or Names).

Fields with the same meaning, but different names, can be assigned the same alias to facilitate queries (e.g. species field, same across all endpoints, in order to be able to search for species). The alias field is used mapped to an ontology entry and currently included in the RDF representation via the owl:sameAs property. The *catalog_reference* table has been extended to hold information on how the feature was generated (ot:hasSource RDF property). The RDF representation of features complies with the ot:Feature class from the OpenTox ontology.

Feature/value pairs belonging to the same experiment are organized in feature tuples, which corresponds to instances of the ot:DataEntry class from the OpenTox ontology.

Templates (tables *template* and *template_def*) allow organizing features in groups. Dataset columns, as well as a model's independent, predicted and observed variables are organized as entries in the templates table. They are automatically created on dataset upload or when applying an algorithm or a model to a dataset.

Templates themselves can be organized hierarchically with the help of the table *dictionary*. The database is distributed with a set of default templates, including top level templates for Endpoints, Identifiers, Datasets and Descriptors as well as a number of endpoints, according to the ECHA endpoints classification[36]. The *ontology* convenience view combines the templates with their hierarchical organisation. This view is currently not used and the corresponding functionality is provided by the OpenTox ontology service.

A full SQL dump of the final OpenTox database structure is provided in Appendix B: OpenTox final database structure.

2. Datasets included in the final database

2.1 Selection of data sources for inclusion

Chemical structures and data from various sources have been imported into the Ambit database and made available via the OpenTox REST API interface[37]. The data sources have been selected within publicly available data sources, providing high-quality structural and/or toxicological data:

- ECHA list of pre-registered substances[38] – The original version of this list was created by ECHA on the basis of information submitted by companies before December 1, 2008. Since then, ECHA has done some tidying up – for example, reconciling the numbers and names of chemicals, putting duplications together and making deletions where companies have requested them. According to ECHA, "some 14,000 of the entries on the list could contain mistakes as far as we can see – sometimes this might be because the name and number don't match; because of language difficulties; because a mixture of substances was given rather than a single substance; or simply because, for whatever reason, a substance doesn't have a number." We have decided to include this data source regardless of its

[36] guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?%20vers=20_08_08 accessed on May 27, 2011

[37] www.opentox.org/dev/apis/api-1.1 accessed on May 27, 2011

[38] apps.echa.europa.eu/preregistered/pre-registered-sub.aspx accessed on May 27, 2011

relatively high error rate (estimated at around 10%), because of its paramount importance and relevance in a REACH[39] context;

- Chemical Identifier Resolver[40] – A REST web service evolved from the Chemical Structure Lookup Service[41] and provided by the NCI/CADD group[42] of the US National Cancer Institute[43]. This service works as a resolver for different chemical structure identifiers and allows one to convert a given structure identifier into another representation or structure identifier. The Chemical Structure DataBase (CSDB) running behind the Chemical Identifier Resolver represents an aggregated collection of over 150 small-molecule databases totalling 103.5 million structure records;
- ChemIDplus[44] – a free, web-based search system that provides access to structure and nomenclature authority files used for the identification of chemical substances cited in the US National Library of Medicine (NLM)[45] databases, including the TOXNET® system[46]. ChemIDplus also provides structure searching and direct links to many biomedical resources at NLM and on the Internet for chemicals of interest. Currently the database contains 390,983 chemical records, of which 299,707 include chemical structures and molecular weight, 139,354 have toxicity data, 25,461 have physical property data, and is searchable by Name, Synonym, CAS Registry Number, Molecular Formula, Classification Code, Locator Code, Structure, Toxicity, and/or Physical properties;
- OPSIN[47] – an open source, freely available, algorithm (Open Parser for Systematic IUPAC Nomenclature, OPSIN) that interprets the majority of organic chemical nomenclature in a fast and precise manner. We have used it to automatically convert the chemical compound names included in the above-mentioned ECHA list of pre-registered substances into chemical structures;
- ChemDraw[48]– a molecule editor developed by the cheminformatics company CambridgeSoft[49]. We have used its "Convert Name to Structure" functionality for manual conversion and further checking of some of the chemical compound names included in the above-mentioned ECHA list of pre-registered substances;
- JRC PRS list[50] – Contains structures for a subset (80,410) of the above-mentioned ECHA list of pre-registered substances, generated by the Computational Toxicology Group, Joint Research Centre,

[39] echa.europa.eu/reach_en.asp accessed on May 27, 2011

[40] cactus.nci.nih.gov/chemical/structure accessed on May 27, 2011

[41] cactus.nci.nih.gov/cgi-bin/lookup/search accessed on May 27, 2011

[42] cactus.nci.nih.gov accessed on May 27, 2011

[43] www.cancer.gov accessed on May 27, 2011

[44] chem.sis.nlm.nih.gov/chemidplus accessed on May 27, 2011

[45] www.nlm.nih.gov accessed on May 27, 2011

[46] toxnet.nlm.nih.gov accessed on May 27, 2011

[47] Daniel M. Lowe, Peter T. Corbett, Peter Murray-Rust, and Robert C. Glen. Chemical Name to Structure: OPSIN, an Open Source Solution. Chem. Inf. Model., 2011, 51 (3), pp 739-753 accessed on May 27, 2011

[48] www.cambridgesoft.com/software/ChemDraw accessed on May 27, 2011

[49] www.cambridgesoft.com accessed on May 27, 2011

[50] ecb.jrc.ec.europa.eu/documents/QSAR/INFORMATION_SOURCES/EC_CHEMICAL_INVENTORIES/ accessed on May 27, 2011

through the ACD/Name to structure Batch[51] software, which converts systematic and non-systematic chemical names of general organic and selected biochemical and inorganic compounds into structures;

- ISSCAN[52] – Istituto Superiore di Sanità[53], "CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA". This database originates from the experience of researchers of the Environment and Primary Prevention Department in the field of structure-activity relationships (SAR), aimed at developing models that theoretically predict the carcinogenicity of chemicals. A portion of the chemicals has been the subject of carcinogenicity classification by various Regulatory Agencies and Scientific Bodies. The database has been specifically designed as an expert decision support tool and includes these carcinogenicity classification "calls" to guide the application of SAR approaches;
- ISSMIC – Istituto Superiore di Sanità; a curated database, containing critically-selected information on chemical compounds tested with the *in vivo* micronucleus mutagenicity assay in rodents. *In vivo* mutagenicity testing is ranked third (well ahead of rodent carcinogenicity) in animal consuming experimentation. Results in bone marrow cells, peripheral blood cells and splenocytes for male/ female rat/mouse, are reported. The data are collected from publicly available databases (Toxnet, NTP), and from the Leadscope FDA CRADA Toxicity Database. ISSMIC provides both biological calls (species, cells, strains, route of administration) and chemical structures, and is the basis for establishing sound read-across and QSAR risk assessment. The Toxicity Biomarker (TB) term has been introduced to show if the substance is positive/negative in cells studied at different doses.
- ISSSTY – Istituto Superiore di Sanità; contains information on chemical compounds tested with the Salmonella typhimurium (Ames) mutagenicity test. The data are collected from the Chemical Carcinogenesis Research Information System (CCRIS) database. The biological data includes Ames test results for each strain tested, with and/or without metabolic activation. The following metabolic activation systems were considered: rat/mouse/hamster – liver – S9/ microsomes/cytosol. Results in 31 strains (with and without metabolic activation) were included.
- DSSTox[54] – a project of the US EPA's National Center for Computational Toxicology[55], helping to build a public data foundation for improved structure-activity and predictive toxicology capabilities. The DSSTox website provides a public forum for publishing downloadable, structure-searchable, standardized chemical structure files associated with toxicity data. In particular, we have included in the final OpenTox database the following DSSTox datasets:
    - CPDBAS[56] – Carcinogenic Potency Database Summary Tables – All Species. The CPDB Summary Tables list summarized results for experiments on 1547 substances in the Carcinogenic Potency Database (CPDB). These Summary Tables report the strongest evidence of carcinogenicity for each chemical, in each sex/species and represent one of many possible summarizations of the data in the CPDB. The CPDB includes detailed results and analyses of 6540 chronic, long-term carcinogenesis bioassays reported in 1513 papers in the general

[51] http://www.acdlabs.com/products/draw_nom/nom/name/ accessed on May 27, 2011

[52] www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7 accessed on May 27, 2011

[53] www.iss.it/chis/?lang=2 accessed on May 27, 2011

[54] www.epa.gov/ncct/dsstox accessed on May 27, 2011

[55] www.epa.gov/ncct accessed on May 27, 2011

[56] www.epa.gov/ncct/dsstox/sdf_cpdbas.html accessed on May 27, 2011

literature and 452 Technical Reports of the US National Cancer Institute/National Toxicology Program[57];

- DBPCAN[58] – EPA Water Disinfection By-Products with Carcinogenicity Estimates. The DBPCAN data file, derived from data published by Woo et. al, 2002[59], contains predicted estimates of carcinogenic potential for 209 chemicals detected in finished drinking water samples having undergone water disinfection treatment;
- EPAFHM[60] – EPA Fathead Minnow Acute Toxicity. The EPA Fathead Minnow Acute Toxicity database was generated by the U.S. EPA Mid-Continental Ecology Division (MED)[61] for the purpose of developing an expert system to predict acute toxicity from chemical structure based on mode of action considerations. Hence, an important and unusual characteristic of this toxicity database is that the 617 tested industrial organic chemicals were expressly chosen to serve as a useful training set for development of predictive quantitative structure-activity relationships (QSARs). A second valuable aspect of this database, from a QSAR modelling perspective, is the inclusion of general mode-of-action (MOA)[62] classifications of acute toxicity response for individual chemicals derived from study results. These MOA assignments are biologically-based classifications, allowing definition of chemical similarity based upon biological activity instead of organic chemistry functional classes as most commonly employed in QSAR studies. MOA classifications should strengthen the scientific basis for construction of individual QSARs;
- FDAMDD[63] – FDA Maximum (Recommended) Daily Dose. The US Food and Drug Administration (FDA) Center for Drug Evaluation and Research[64], Office of Pharmaceutical Science, Informatics and Computational Safety Analysis Staff's Maximum Recommended Daily Dose (FDAMDD) database contains values for over 1200 pharmaceuticals listed in Martindale: The Extra Pharmacopoeia (1973, 1983, and 1993) and The Physicians' Desk Reference (1995 and 1999)[65]. Most of the maximum recommended daily dose (MRDD) values in the database were determined from pharmaceutical clinical trials that employed an oral route of exposure and daily treatments, usually for 3-12 months;
- HPVCSI[66] – EPA High Production Volume (HPV) Challenge Program; consolidates three historical EPA HPV Chemical Lists[67] associated with multiple public notices of the HPV voluntary chemical adoption program;

---

[57] ntp.niehs.nih.gov accessed on May 27, 2011

[58] www.epa.gov/ncct/dsstox/sdf_dbpcan.html accessed on May 27, 2011

[59] Woo, Y.T., D. Lai, J.L. McLain, M.K. Manibusan, and V. Dellarco (2002) Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products, Environ. Health Perspect.,110 Suppl 1: 75-87. www.epa.gov/ncct/dsstox/Citations/Woo_et_al.2002_DBP_SAR.pdf accessed on May 27, 2011

[60] www.epa.gov/ncct/dsstox/sdf_epafhm.html accessed on May 27, 2011

[61] www.epa.gov/med accessed on May 27, 2011

[62] www.epa.gov/ncct/dsstox/CentralFieldDef.html#MOA accessed on May 27, 2011

[63] www.epa.gov/ncct/dsstox/sdf_fdamdd.html accessed on May 27, 2011

[64] www.fda.gov/Drugs/default.htm accessed on May 27, 2011

[65] en.wikipedia.org/wiki/Martindale:_The_Extra_Pharmacopoeia accessed on May 27, 2011

[66] www.epa.gov/ncct/dsstox/sdf_hpvcsi.html accessed on May 27, 2011

- HPVISD[68] – EPA High Production Volume Information System (HPV-IS) Data; HPVIS consists of basic hazard (toxicity) and environmental fate information on HPV chemicals that can be used by environmental managers, public decision-makers, and others in their own health and environmental protection activities. HPVIS submissions contain data on up to 50 endpoints organized into the following four disciplines: physical/chemical properties (e.g., melting point, vapour pressure) environmental fate and pathways (e.g., biodegradation, stability in soil) ecotoxicity (e.g., fish toxicity, toxicity to aquatic plants) and mammalian health effects (e.g., reproductive toxicity, developmental toxicity);
- KIERBL[69] – EPA Estrogen Receptor Ki Binding Study (Laws et al.). This study was conducted by US EPA researchers to evaluate the validity of the rat uterine cytosolic (RUC) estrogen receptor (ER) competitive binding assay for use in the Endocrine Disruption Screening Program (EDSP)[70]. The assay measures the ability of radiolabeled 17-beta-estradiol (^{3}H-E2) to bind with RUC ER in the presence of increasing concentrations of a test chemical. The data files include all published IC50 and Ki experimental results for the 50 chemicals included in Laws et al., 2006[71], (denoted Group 1), as well as previously unpublished results for an additional 228 structurally diverse TSCA chemicals[72] for which no ER binding was observed (denoted Group 2);
- IRISTR[73] – EPA Integrated Risk Information System (IRIS) Toxicity Review Data. IRIS is a database of human health effects that may result from exposure to various substances found in the environment. IRIS chemical files contain descriptive and quantitative information in the following categories:
  - Oral reference doses and inhalation reference concentrations (RfDs and RfCs, respectively) for chronic non-carcinogenic health effects;
  - Hazard identification, oral slope factors, and oral and inhalation unit risks for carcinogenic effects;
- NCTRER[74] – FDA National Center for Toxicological Research Estrogen Receptor Binding. Consists of 232 chemicals (131 active and 101 inactive) selected *a priori* based on structural characteristics and tested in a well-validated and standardized *in vitro* rat uterine cytosol ER competitive-binding assay [Blair et al. 2000; Branham et al., 2002]. The database is a structurally diverse set of natural, synthetic, and environmental estrogens covering most

---

[67] www.epa.gov/chemrtk/pubs/update/hpvchmlt.htm accessed on May 27, 2011

[68] www.epa.gov/ncct/dsstox/sdf_hpvisd.html accessed on May 27, 2011

[69] www.epa.gov/ncct/dsstox/sdf_kierbl.html accessed on May 27, 2011

[70] www.epa.gov/scipoly/oscpendo accessed on May 27, 2011

[71] Laws SC, Yavanhxay S, Copper RL, Eldridge JC. 2006. Nature of the binding interaction for 50 structurally diverse chemicals with rat estrogen receptors. Toxicological Sciences.94(1), 46-56; doi:10.1093/toxsci/kfl092.toxsci.oxfordjournals.org/cgi/reprint/kfl092?ijkey=qOBz5uEzojIv0zy&keytype=ref accessed on May 27, 2011

[72] www.epa.gov/lawsregs/laws/tsca.html accessed on May 27, 2011

[73] www.epa.gov/ncct/dsstox/sdf_iristr.html accessed on May 27, 2011

[74] www.epa.gov/ncct/dsstox/sdf_nctrer.html accessed on May 27, 2011

known estrogenic classes and spanning a wide range of biological activity. It represents the largest published ER binding database of same-assay results generated in a single laboratory;

- NTPBSI[75] – National Toxicology Program Bioassay On-line Database. Contains results collected on approximately 300 toxicity studies from shorter duration tests and from more than 2000 genetic toxicity studies, some of which include both *in vitro* and *in vivo* tests;
- NTPHTS[76] – National Toxicology Program High Throughput Screening Project. Contains HTS data for a set of 1408 chemicals from NTP inventories. Assays were selected based on their potential to be informative of animal bioassay results and relevant to human health risk assessments;

- TOXCST[77] – Screening Chemicals to Predict Toxicity Faster and Better. Includes over 500 state-of-the-art rapid tests (called high-throughput assays) that are screening 309 environmental chemicals for potential toxicity; inclusion of duplicates and triplicates yields 320 entries in total. Data is further subdivided in the following 9 datasets: TOXCST_ACEA, TOXCST_Attagene, TOXCST_BioSeek, TOXCST_Cellumen, TOXCST_CellzDirect, TOXCST_Gentronix, TOXCST_NCGC, TOXCST_Novascreen, TOXCST_Solidus. These are complemented by TOXCST_ToxRefDB, that contains over 30 years' worth of existing toxicity data on these chemicals, collected primarily through animal studies;
- TXCST2[78] – Screening Chemicals to Predict Toxicity Faster and Better. A collection of 960 chemicals from a broad range of sources including industrial and consumer products, food additives and drugs that never made it to the market. These chemicals are currently subject to HTS to evaluate the predictive toxicity signatures developed in Phase I (see TOXCST dataset);
- ECETOC skin irritation[79] – 176 chemicals for which comprehensive rabbit skin irritation/corrosion data are available. The chemicals represent a range of chemical classes (acids, acrylates/methacrylates, alcohols, aldehydes, alkalis, amides, amines, brominated derivatives, chlorinated solvents, esters, ethers, fatty acids and mixtures, fragrance oils, halogenated aromatics, hydrocarbons (unsaturated), inorganics, ketones, nitriles, phenolic derivatives, S-containing compounds, soaps/surfactants, triglycerides) and different degrees of irritancy;
- Skin sensitisation (LLNA)[80] – a database that comprises local lymph node assay (LLNA) data on 211 individual chemicals, encompassing both the chemical and biological diversity of known chemical allergens. To cover the range of relative allergenic potencies, the data set includes data on 13 extreme, 21 strong, 69 moderate, and 66 weak contact allergens, classified according to each allergen's mathematically estimated concentration required to induce a threefold stimulation index. In addition, there are also 42 chemicals that are considered to be non-sensitizers. In terms of chemical diversity,

[75] http://www.epa.gov/ncct/dsstox/sdf_ntpbsi.html accessed on May 27, 2011

[76] http://www.epa.gov/ncct/dsstox/sdf_ntphts.html accessed on May 27, 2011

[77] www.epa.gov/ncct/toxcast/ accessed on May 27, 2011

[78] ToxCast Phase II Chemicals http://www.epa.gov/ncct/toxcast/files/TXCST2_v1a_960_23Feb2011.zip accessed on August 21, 2011

[79] ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) accessed on May 27, 2011

[80] Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz G, Basketter DA. (2005). Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives. Dermatitis 16(4): 157-202. accessed on May 27, 2011

the database contains data pertaining to the chemical classes represented by aldehydes, ketones, aromatic amines, quinones, and acrylates, as well as compounds that have different reactivity mechanisms. The list of chemicals contained in the data set represents both the chemical and biological diversity that is known to exist for chemical allergens and non-allergens;

- Skin sensitisation (LLNA – second compilation)[81] – derived from previously conducted LLNA studies involving an additional 108 chemicals;
- Bioconcentration factor (BCF) Gold Standard Database[82] – a database holding peer-reviewed high-quality BCF data that is a valuable resource for development of alternative tests;
- Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way[83] – pKa values for 185 chemicals;
- Benchmark Data Set for In Silico Prediction of Ames Mutagenicity[84] – Ames mutagenicity data set comprising 6512 non-confidential compounds together with their biological activity;
- Bursi AMES Toxicity Dataset[85] – 4337 Compounds with Mutagenicity (AMES) classification (2401 mutagens and 1936 non-mutagens);
- EpiSuite[86] – 14 datasets with various physico-chemical and environmental fate properties (EPI_AOP, EPI_BCF, EPI_BioHC, EPI_Biowin, EPI_Boil_Pt, EPI_Henry, EPI_KM, EPI_KOA, EPI_Kowwin, EPI_Melt_Pt, EPI_PCKOC, EPI_VP, EPI_WaterFrag, EPI_Wskowwin);
- PubChem Structures & Assays[87] – 473,965 structures from PubChem with assay data included as properties, suitable for building QSAR or other types of models;
- Leadscope[88] – 12 datasets, accessible exclusively through Authorisation Policies, e.g., restricted to OpenTox partners (Leadscope_carc_level_2, Leadscope_ccris_genetox, Leadscope_cder_chronic, Leadscope_cder_genetox, Leadscope_cder_repro_dev, Leadscope_cfsan_acute, Leadscope_cfsan_chronic, Leadscope_cfsan_genetox, Leadscope_cfsan_repro_dev, Leadscope_fda_marketed_drugs, Leadscope_genetox_level_2, Leadscope_ntp_genetox);
- Pharmatrope[89] – AERS hepatobiliary system data, accessible exclusively through Authorisation Policies, e.g., restricted to OpenTox partners.

2.2 Included datasets details

In the following sub-sections we provide further details for each of the 67 final OpenTox database datasets. Please note that one needs to login to access these datasets. Most of the datasets are publicly accessible

[81] Kern PS, Gerberick GF, Ryan CA, Kimber I, Aptula A, Basketter DA. (2010). Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. Dermatitis 21(1): 8-32. accessed on May 27, 2011

[82] ambit.sourceforge.net/euras accessed on May 27, 2011

[83] Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way accessed on May 27, 2011

[84] Benchmark Data Set for In Silico Prediction of Ames Mutagenicity accessed on May 27, 2011

[85] Bursi AMES Toxicity Dataset accessed on May 27, 2011

[86] EpiSuite data accessed on May 27, 2011

[87] Downloadable Structure Files of PubChem Compounds accessed on May 27, 2011

[88] http://www.leadscope.com/ accessed on May 27, 2011

[89] http://www.pharmatrope.com/ accessed on May 27, 2011

through the OpenTox *guest* account (username: guest, password: guest). Leadscope and Pharmatrope datasets are currently accessible only to OpenTox partners.

2.2.1 ECHA list of pre-registered substances

- Name: ECHA list of pre-registered substances (20090327)
- URI:https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/1?pagesize=10&page=0
- Number of compounds: 143,835
- Number of empty structures: 143,835
- CAS numbers OK: 118,284
- CAS numbers errors: 2
- EINECS numbers OK: 143,654
- EINECS numbers errors: 181

2.2.2 Chemical Identifier Resolver

- Name: Chemical Identifier Resolver (20110212)
- URI:https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/5?pagesize=10&page=0
- Number of compounds: 72,985
- Number of empty structures: 0
- CAS numbers OK: 72,985
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.3 ChemIDplus

- Name: ChemIDplus (20110503)
- URI:https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/3?pagesize=10&page=0
- Number of compounds: 80,468
- Number of empty structures: 0
- CAS numbers OK: 80,468
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.4 ChemDraw

- Name: ChemDraw (20110505)
- URI:https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/7?pagesize=10&page=0
- Number of compounds: 22,519
- Number of empty structures: 0
- CAS numbers OK: 0

- CAS numbers errors: 0
- EINECS numbers OK: 22,519
- EINECS numbers errors: 0

2.2.5 JRC PRS list

- Name: ECBPRS (20090917)
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6483?pagesize=10&page=0
- Number of compounds: 80,410
- Number of empty structures: 0
- CAS numbers OK: 80,410
- CAS numbers errors: 0
- EINECS numbers OK: 70,749
- EINECS numbers errors: 0

2.2.6 ISSCAN

- Name: ISSCAN
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/31?pagesize=10&page=0
- Number of compounds: 1150
- Number of empty structures: 12
- CAS numbers OK: 1139
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.7 ISSMIC

- Name: ISSMIC
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/33?pagesize=10&page=0
- Number of compounds: 151
- Number of empty structures: 6
- CAS numbers OK: 150
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.8 ISSSTY

- Name: ISSSTY
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/35?pagesize=10&page=0
- Number of compounds: 223

- Number of empty structures: 1
- CAS numbers OK:0
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.9 CPDBAS

- Name: CPDBAS
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/9?pagesize=10&page=0
- Number of compounds: 1547
- Number of empty structures: 39
- CAS numbers OK: 1528
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.10 DBPCAN

- Name: DBPCAN
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/11?pagesize=10&page=0
- Number of compounds: 209
- Number of empty structures: 0
- CAS numbers OK: 179
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.11 EPAFHM

- Name: EPAFHM
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/13?pagesize=10&page=0
- Number of compounds: 617
- Number of empty structures: 0
- CAS numbers OK: 617
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.12 FDAMDD

- Name: FDAMDD

- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/15?pagesize=10&page=0
- Number of compounds: 1216
- Number of empty structures: 0
- CAS numbers OK: 1216
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.13 HPVCSI

- Name: HPVCSI
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/17?pagesize=10&page=0
- Number of compounds: 3548
- Number of empty structures: 1068
- CAS numbers OK: 3548
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.14 HPVISD

- Name: HPVISD
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/19?pagesize=10&page=0
- Number of compounds: 1006
- Number of empty structures: 302
- CAS numbers OK: 1006
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.15 IRISTR

- Name: IRISTR
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/21?pagesize=10&page=0
- Number of compounds: 544
- Number of empty structures: 5
- CAS numbers OK: 536
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.16 KIERBL

- Name: KIERBL
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/23?pagesize=10&page=0
- Number of compounds: 278
- Number of empty structures: 0
- CAS numbers OK: 278
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.17 NCTRER

- Name: NCTRER
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/25?pagesize=10&page=0
- Number of compounds:232
- Number of empty structures: 0
- CAS numbers OK: 227
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.18 NTPBSI

- Name: NTPBSI
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/27?pagesize=10&page=0
- Number of compounds: 2330
- Number of empty structures: 151
- CAS numbers OK: 2218
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.19 NTPHTS

- Name: NTPHTS
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/29?pagesize=10&page=0
- Number of compounds: 1408
- Number of empty structures: 14
- CAS numbers OK: 1407
- CAS numbers errors: 0

- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.20 TOXCST

- Name: TOXCST
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/37?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.21 TOXCST_ACEA

- Name: TOXCST_ACEA
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/83?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.22 TOXCST_Attagene

- Name: TOXCST_Attagene
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/723?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.23 TOXCST_BioSeek

- Name: TOXCST_BioSeek
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/1363?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0

- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.24 TOXCST_Cellumen

- Name: TOXCST_Cellumen
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/2003?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.25 TOXCST_CellzDirect

- Name: TOXCST_CellzDirect
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/2643?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.26 TOXCST_Gentronix

- Name: TOXCST_Gentronix
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/3283?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.27 TOXCST_NCGC

- Name: TOXCST_NCGC
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/3923?pagesize=10&page=0

- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.28 TOXCST_Novascreen

- Name: TOXCST_Novascreen
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/4563?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.29 TOXCST_Solidus

- Name: TOXCST_Solidus
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/5203?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.30 TOXCST_ToxRefDB

- Name: TOXCST_ToxRefDB
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/5843?pagesize=10&page=0
- Number of compounds: 320
- Number of empty structures: 0
- CAS numbers OK: 320
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.31 TXCST2

- Name: TXCST2
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/39?pagesize=10&page=0
- Number of compounds: 960
- Number of empty structures: 0
- CAS numbers OK: 927
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.32 ECETOC skin irritation

- Name: ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/41?pagesize=10&page=0
- Number of compounds: 176
- Number of empty structures: 1
- CAS numbers OK: 172
- CAS numbers errors: 4
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.33 Skin sensitisation (LLNA)

- Name: Local Lymph Node Data for the Evaluation of Skin Sensitization – Compilation of historical data (Dermatitis Vol 16 No 4 2005)
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/43?pagesize=10&page=0
- Number of compounds: 209
- Number of empty structures: 0
- CAS numbers OK: 203
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.34 Skin sensitisation (LLNA) – second compilation

- Name: Local Lymph Node Data for the Evaluation of Skin Sensitization – Second compilation (Dermatitis Vol 21 No 1 2010)
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/45?pagesize=10&page=0
- Number of compounds: 108
- Number of empty structures: 0

- CAS numbers OK: 99
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.35 Bioconcentration factor (BCF) gold standard database

- Name: Bioconcentration factor (BCF) Gold Standard Database
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/47?pagesize=10&page=0
- Number of compounds: 1130
- Number of empty structures: 1130
- CAS numbers OK: 1122
- CAS numbers errors: 7
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.36 Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way

- Name: Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/49?pagesize=10&page=0
- Number of compounds: 185
- Number of empty structures:0
- CAS numbers OK: 0
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.37 Benchmark Data Set for In Silico Prediction of Ames Mutagenicity

- Name: Benchmark Data Set for In Silico Prediction of Ames Mutagenicity
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/51?pagesize=10&page=0
- Number of compounds: 6512
- Number of empty structures: 0
- CAS numbers OK: 6197
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.38 Bursi AMES Toxicity Dataset

- Name: Bursi AMES Toxicity Dataset
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/53?pagesize=10&page=0

- Number of compounds: 4337
- Number of empty structures: 0
- CAS numbers OK: 4337
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.39 EPI_AOP

- Name: EPI_AOP
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/55?pagesize=10&page=0
- Number of compounds: 818
- Number of empty structures: 0
- CAS numbers OK: 810
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.40 EPI_BCF

- Name: EPI_BCF
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/57?pagesize=10&page=0
- Number of compounds: 685
- Number of empty structures: 0
- CAS numbers OK: 675
- CAS numbers errors: 1
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.41 EPI_BioHC

- Name: EPI_BioHC
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/59?pagesize=10&page=0
- Number of compounds: 175
- Number of empty structures: 0
- CAS numbers OK: 170
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.42 EPI_Biowin

- Name: EPI_Biowin
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/61?pagesize=10&page=0
- Number of compounds: 1263
- Number of empty structures: 0
- CAS numbers OK: 1215
- CAS numbers errors: 1
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.43 EPI_Boil_Pt

- Name: EPI_Boil_Pt
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/63?pagesize=10&page=0
- Number of compounds: 5890
- Number of empty structures: 1
- CAS numbers OK: 5859
- CAS numbers errors: 1
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.44 EPI_Henry

- Name: EPI_Henry
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/65?pagesize=10&page=0
- Number of compounds: 1829
- Number of empty structures: 0
- CAS numbers OK: 1821
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.45 EPI_KM

- Name: EPI_KM
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/67?pagesize=10&page=0
- Number of compounds: 631
- Number of empty structures: 0
- CAS numbers OK: 602
- CAS numbers errors: 0

- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.46 EPI_KOA

- Name: EPI_KOA
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/69?pagesize=10&page=0
- Number of compounds: 308
- Number of empty structures: 0
- CAS numbers OK: 259
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.47 EPI_Kowwin

- Name: EPI_Kowwin
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/71?pagesize=10&page=0
- Number of compounds: 15,809
- Number of empty structures: 0
- CAS numbers OK: 12,162
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.48 EPI_Melt_Pt

- Name: EPI_Melt_Pt
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/73?pagesize=10&page=0
- Number of compounds: 10,051
- Number of empty structures: 5
- CAS numbers OK: 10,028
- CAS numbers errors: 1
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.49 EPI_PCKOC

- Name: EPI_PCKOC
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/75?pagesize=10&page=0
- Number of compounds: 788
- Number of empty structures: 2

- CAS numbers OK: 770
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.50 EPI_VP

- Name: EPI_VP
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/77?pagesize=10&page=0
- Number of compounds: 3037
- Number of empty structures: 1
- CAS numbers OK: 3031
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.51 EPI_WaterFrag

- Name: EPI_WaterFrag
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/79?pagesize=10&page=0
- Number of compounds: 5764
- Number of empty structures: 3
- CAS numbers OK: 5753
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.52 EPI_Wskowwin

- Name: EPI_Wskowwin
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/81?pagesize=10&page=0
- Number of compounds: 2348
- Number of empty structures:3
- CAS numbers OK: 2346
- CAS numbers errors: 1
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.53 Chemical Name to Structure Converter (OPSIN)

- Name: NAME2STRUCTURE
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6487?pagesize=10&page=0

- Number of compounds: 70,646
- Number of empty structures: 16
- CAS numbers OK: 0
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.54 PubChem Structures + Assays

- Name: PubChem Structures + Assays
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6489?pagesize=10&page=0
- Number of compounds: 473,965
- Number of empty structures: 45
- CAS numbers OK: 1275
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.55 Leadscope_carc_level_2

- Name: Leadscope_carc_level_2
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6490?pagesize=10&page=0
- Number of compounds: 2988
- Number of empty structures: 141
- CAS numbers OK: 600
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.56 Leadscope_ccris_genetox

- Name: Leadscope_ccris_genetox
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6492?pagesize=10&page=0
- Number of compounds: 8001
- Number of empty structures: 559
- CAS numbers OK: 1410
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.57 Leadscope_cder_chronic

- Name: Leadscope_cder_chronic
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6494?pagesize=10&page=0
- Number of compounds: 121
- Number of empty structures: 4
- CAS numbers OK: 2
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.58 Leadscope_cder_genetox

- Name: Leadscope_cder_genetox
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6496?pagesize=10&page=0
- Number of compounds: 336
- Number of empty structures: 242
- CAS numbers OK: 12
- CAS numbers errors: 1
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.59 Leadscope_cder_repro_dev

- Name: Leadscope_cder_repro_dev
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6498?pagesize=10&page=0
- Number of compounds: 58
- Number of empty structures: 0
- CAS numbers OK: 2
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.60 Leadscope_cfsan_acute

- Name: Leadscope_cfsan_acute
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6500?pagesize=10&page=0
- Number of compounds: 1070
- Number of empty structures: 100
- CAS numbers OK: 690
- CAS numbers errors: 0

- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.61 Leadscope_cfsan_chronic

- Name: Leadscope_cfsan_chronic
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6502?pagesize=10&page=0
- Number of compounds: 655
- Number of empty structures: 106
- CAS numbers OK: 307
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.62 Leadscope_cfsan_genetox

- Name: Leadscope_cfsan_genetox
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6504?pagesize=10&page=0
- Number of compounds: 696
- Number of empty structures: 130
- CAS numbers OK: 247
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.63 Leadscope_cfsan_repro_dev

- Name: Leadscope_cfsan_repro_dev
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6506?pagesize=10&page=0
- Number of compounds: 312
- Number of empty structures: 47
- CAS numbers OK: 135
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.64 Leadscope_fda_marketed_drugs

- Name: Leadscope_fda_marketed_drugs
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6508?pagesize=10&page=0
- Number of compounds: 6637
- Number of empty structures: 321

- CAS numbers OK: 451
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.65 Leadscope_genetox_level_2

- Name: Leadscope_genetox_level_2
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6510?pagesize=10&page=0
- Number of compounds: 10,155
- Number of empty structures: 748
- CAS numbers OK: 1723
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.66 Leadscope_ntp_genetox

- Name: Leadscope_ntp_genetox
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6512?pagesize=10&page=0
- Number of compounds: 2128
- Number of empty structures: 71
- CAS numbers OK: 956
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.67 Pharmatrope_AERS_hepatobiliary_system

- Name: Pharmatrope_AERS_hepatobiliary_system
- URI: https://ambit.uni-plovdiv.bg:8443/ambit2/dataset/6514?pagesize=10&page=0
- Number of compounds: 1274
- Number of empty structures: 0
- CAS numbers OK: 0
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.3 Quality Labels Overview

Automatic classification has been applied to all newly imported structures. The results from this exercise are presented in Table 2. The first column indicates whether we have a Consensus, Majority, Unconfirmed or Ambiguous label, the second column provides further details for each of these categories, while the last

column lists the number of chemical structures in the given category. When a 'Consensus' label is assigned, the Details column lists the number of contributing independent sources and measures how strong this consensus is (e.g. 2, 3, 4, etc.). In the case of 'Majority' label the Details column provides insight on the number of alternatives and the distribution of the votes from independent sources for each of these alternatives (e.g. 1:1:4, 2:3, etc.). The label 'Unconfirmed' is assigned when the available data comes from a single source. And finally for the 'Ambiguous' label the Details column lists the number of alternatives and the distribution of the votes from independent sources (e.g. 1:1, 2:2, 1:1:1:1, etc.).

Table 2 Final OpenTox database automatic classification results distribution

Label	Details	Number of Chemicals
Consensus	46	1
Consensus	42	1
Consensus	40	1
Consensus	39	3
Consensus	38	2
Consensus	37	1
Consensus	36	10
Consensus	35	7
Consensus	34	10
Consensus	33	14
Consensus	32	16
Consensus	31	23
Consensus	30	20
Consensus	29	28
Consensus	28	24
Consensus	27	49
Consensus	26	46
Consensus	25	49
Consensus	24	65
Consensus	23	51
Consensus	22	79
Consensus	21	72
Consensus	20	80
Consensus	19	93
Consensus	18	115
Consensus	17	157
Consensus	16	128
Consensus	15	177
Consensus	14	206
Consensus	13	310
Consensus	12	329
Consensus	11	452
Consensus	10	615
Consensus	9	780
Consensus	8	1208
Consensus	7	2080
Consensus	6	3552
Consensus	5	10,908
Consensus	4	26,390
Consensus	3	10,311
Consensus	2	12,272
Majority	10:2	11
Majority	10:3	11
Majority	10:4	6
Majority	10:5	5
Majority	10:6	2
Majority	10:7	1
Majority	10:8	1
Majority	11:14	1
Majority	11:2	8
Majority	11:3	15
Majority	11:4	13
Majority	11:5	5
Majority	11:6	2
Majority	11:7	2
Majority	12:2	4
Majority	12:2:4	1
Majority	12:3	13
Majority	12:4	9
Majority	12:5	3
Majority	12:6	2
Majority	13:2	4
Majority	13:3	9
Majority	13:4	16
Majority	13:5	3
Majority	13:6	3

Label	Details	Number of Chemicals
Majority	14:3	7
Majority	14:4	7
Majority	14:5	9
Majority	14:6	4
Majority	14:8	1
Majority	15:2	1
Majority	15:3	4
Majority	15:4	3
Majority	15:5	1
Majority	15:6	2
Majority	16:3	6
Majority	16:4	4
Majority	16:5	2
Majority	17:2	3
Majority	17:3	4
Majority	17:4	3
Majority	17:5	2
Majority	17:6	1
Majority	17:8	1
Majority	18:3	2
Majority	18:4	1
Majority	18:5	3
Majority	18:6	1
Majority	19:2	1
Majority	19:3	2
Majority	19:4	4
Majority	19:5	4
Majority	1:10	38
Majority	1:10:2	1
Majority	1:10:3	1
Majority	1:11	31
Majority	1:11:2	2
Majority	1:11:4	1
Majority	1:12	26
Majority	1:12:4	2
Majority	1:13	19
Majority	1:13:3	1
Majority	1:14	11
Majority	1:14:6	1
Majority	1:15	8
Majority	1:15:2	1

Label	Details	Number of Chemicals
Majority	1:16	8
Majority	1:16:5	1
Majority	1:17	3
Majority	1:17:9	1
Majority	1:18	8
Majority	1:18:6	1
Majority	1:19	2
Majority	1:1:10	2
Majority	1:1:11	3
Majority	1:1:11:6	1
Majority	1:1:12	2
Majority	1:1:13	1
Majority	1:1:13:6	1
Majority	1:1:16	1
Majority	1:1:16:7	1
Majority	1:1:1:16	1
Majority	1:1:1:17	1
Majority	1:1:1:1:3	2
Majority	1:1:1:2	28
Majority	1:1:1:2:4	1
Majority	1:1:1:3	13
Majority	1:1:1:4	1
Majority	1:1:1:5	1
Majority	1:1:1:6	1
Majority	1:1:1:7	1
Majority	1:1:2	1178
Majority	1:1:24	1
Majority	1:1:2:3	3
Majority	1:1:2:4	1
Majority	1:1:3	274
Majority	1:1:3:4	1
Majority	1:1:3:7	1
Majority	1:1:3:8	1
Majority	1:1:4	51
Majority	1:1:4:5	1
Majority	1:1:4:8	1
Majority	1:1:5	34
Majority	1:1:6	8
Majority	1:1:7	10
Majority	1:1:8	7
Majority	1:1:9	2

Label	Details	Number of Chemicals
Majority	1:2	3690
Majority	1:20	1
Majority	1:21	3
Majority	1:22	1
Majority	1:23	1
Majority	1:24	1
Majority	1:27	2
Majority	1:29	1
Majority	1:2:2:5	2
Majority	1:2:3	35
Majority	1:2:3:3:4	1
Majority	1:2:4	19
Majority	1:2:5	13
Majority	1:2:6	12
Majority	1:2:7	4
Majority	1:2:8	2
Majority	1:3	6007
Majority	1:30	2
Majority	1:3:4	17
Majority	1:3:5	8
Majority	1:3:6	3
Majority	1:3:7	8
Majority	1:3:8	5
Majority	1:3:9	1
Majority	1:4	2003
Majority	1:4:4:8	1
Majority	1:4:5	2
Majority	1:4:6	2
Majority	1:4:7	2
Majority	1:4:8	2
Majority	1:4:9	2
Majority	1:5	506
Majority	1:5:7	3
Majority	1:5:8	1
Majority	1:5:9	2
Majority	1:6	250
Majority	1:6:7	2
Majority	1:7	164
Majority	1:8	73
Majority	1:8:9	1
Majority	1:9	67

Label	Details	Number of Chemicals
Majority	20:3	1
Majority	20:4	3
Majority	20:5	2
Majority	20:6	1
Majority	21:4	3
Majority	21:5	6
Majority	22:4	1
Majority	22:5	4
Majority	22:8	1
Majority	23:4	1
Majority	23:5	4
Majority	23:6	1
Majority	24:3	1
Majority	24:4	1
Majority	24:5	1
Majority	25:3	1
Majority	25:4	2
Majority	25:5	1
Majority	26:3	1
Majority	26:4	2
Majority	26:5	2
Majority	27:6	1
Majority	2:20	2
Majority	2:2:3	1
Majority	2:2:4	2
Majority	2:2:5	1
Majority	2:2:6	1
Majority	2:2:7	2
Majority	2:2:8	1
Majority	2:3	516
Majority	2:3:4	2
Majority	2:3:7	2
Majority	2:3:9	1
Majority	2:4	157
Majority	2:4:5	1
Majority	2:4:5:9	1
Majority	2:4:6	1
Majority	2:4:7	1
Majority	2:4:8	1
Majority	2:4:9	2
Majority	2:5	58

Label	Details	Number of Chemicals
Majority	2:5:8	1
Majority	2:6	54
Majority	2:7	30
Majority	2:8	25
Majority	2:9	16
Majority	3:3:6	1
Majority	3:4	102
Majority	3:4:7	1
Majority	3:5	59
Majority	3:5:7	1
Majority	3:6	39
Majority	3:7	27
Majority	3:8	31
Majority	3:9	21
Majority	4:4:5	1
Majority	4:5	29
Majority	4:5:6	1
Majority	4:6	20
Majority	4:7	14
Majority	4:8	15
Majority	4:9	16
Majority	5:6	11
Majority	5:7	9
Majority	5:8	3
Majority	5:9	7
Majority	6:7	6

Label	Details	Number of Chemicals
Majority	6:8	12
Majority	6:9	3
Majority	7:8	2
Unconfirmed	1	498,953
Ambiguous	1:1	2220
Ambiguous	1:1:1	485
Ambiguous	1:1:1:1	79
Ambiguous	1:1:1:1:1	10
Ambiguous	1:1:1:2:2	1
Ambiguous	1:1:2:2	8
Ambiguous	1:1:3:3	2
Ambiguous	1:2:2	93
Ambiguous	1:2:3:3	1
Ambiguous	1:3:3	13
Ambiguous	1:4:4	5
Ambiguous	1:6:6	1
Ambiguous	2:2	1227
Ambiguous	2:2:2	1
Ambiguous	2:2:2:2	1
Ambiguous	2:4:4	1
Ambiguous	3:3	82
Ambiguous	4:4	26
Ambiguous	5:5	6
Ambiguous	6:6	5
Ambiguous	7:7	2

To obtain a better assessment of the quality of various data sources, we have summarized the assigned initial quality labels (according to the mapping between automatic classification and initial quality labels presented above in the Data provenance and quality assurance section).

Table *3* lists the results and highlights the fact that ChemIDplus (3.302% error rate) and JRC PRS list data (error rate 3.087%) seem to provide better quality data when compared to other generic data sources such as the Chemical Identifier Resolver data (6.345% error rate). Most of the DSSTox and EPI datasets prove to have excellent quality (0% or close to 0% error rate).

Table 3 Final OpenTox database quality labels distribution

Dataset	OK	Probably OK	Unknown	Probably ERROR	Probably ERROR%
ChemIDplus (20110503)	59,620	12,694	5497	2657	3.302%
Chemical Identifier Resolver (20110212)	55,470	9719	3164	4631	6.345%
ChemDraw (20110505)	18,147	3542	263	567	2.518%
CPDBAS	1011	372	78	47	3.038%
DBPCAN	150	5	54	0	0%
EPAFHM	510	78	28	1	0.162%
FDAMDD	759	291	148	18	1.480%
HPVCSI	1983	342	145	10	0.282%
HPVISD	603	98	3	0	0%
IRISTR	414	99	23	3	0.551%
KIERBL	224	48	5	1	0.360%
NCTRER	180	17	35	0	0%
NTPBSI	1583	520	48	28	1.202%
NTPHTS	1062	322	1	9	0.639%
ISSCAN	714	265	134	25	2.174%
ISSMIC	105	32	4	4	2.650%
ISSSTY	157	55	5	5	2.242%
TOXCST	253	60	1	0	0%
TXCST2	677	156	121	6	0.625%
ECETOC	163	9	2	1	0.568%
LLNA	125	16	67	1	0.478%
LLNA–2nd compilation	66	12	29	1	0.926%
Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way	50	7	128	0	0%
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity	3687	428	2204	193	2.964%
Bursi AMES Toxicity Dataset	2311	587	1071	368	8.485%
EPI_AOP	713	39	61	5	0.611%
EPI_BCF	547	107	22	9	1.314%
EPI_BioHC	159	0	16	0	0%
EPI_Biowin	1045	157	53	8	0.633%
EPI_Boil_Pt	5124	299	436	30	0.509%
EPI_Henry	1605	203	15	6	0.328%
EPI_KM	511	67	53	0	0%
EPI_KOA	244	15	49	0	0%
EPI_Kowwin	5730	956	8991	132	0.835%
EPI_Melt_Pt	7864	1167	812	203	2.020%
EPI_PCKOC	616	124	36	10	1.269%
EPI_VP	2697	286	42	11	0.362%
EPI_WaterFrag	3980	588	1119	74	1.284%

Dataset	OK	Probably OK	Unknown	Probably ERROR	Probably ERROR%
EPI_Wskowwin	2001	331	6	7	0.298%
ECBPRS	56,079	12,346	9503	2482	3.087%
NAME2STRUCTURE (OPSIN)	49,251	6116	8635	6628	9.382%
PubChem Structures + Assays	7920	1424	464,355	221	0.047%
Leadscope_carc_level_2	1899	285	312	351	11.747%
Leadscope_ccris_genetox	5743	875	196	628	7.849%
Leadscope_cder_chronic	83	20	5	9	7.438%
Leadscope_cder_genetox	75	14	1	4	1.190%
Leadscope_cder_repro_dev	42	13	0	3	5.172%
Leadscope_cfsan_acute	806	58	94	12	1.121%
Leadscope_cfsan_chronic	458	54	26	11	1.679%
Leadscope_cfsan_genetox	486	54	15	11	1.580%
Leadscope_cfsan_repro_dev	211	38	5	11	3.526%
Leadscope_fda_marketed_drugs	2579	465	2595	677	10.200%
Leadscope_genetox_level_2	6970	1033	646	758	7.464%
Leadscope_ntp_genetox	1580	140	8	329	15.461%
Pharmatrope_AERS_hepatobiliary_system	685	98	320	171	13.422%
Entire database	**317,727**	**57,146**	**511,685**	**21,380**	**2.355%**

The relatively high error rate for the name-to-structure conversion through OPSIN might be (partially) due to errors in the chemical names in the original ECHA list of pre-registered substances. A more thorough investigation would be required to check this.

It might be worth recalling that the Unknown class is further subdivided in 'unknown multiple source' (12,735 instances with ambiguous/contradictory information) and 'unknown single source' (498,950 instances, 93% of which are due to the PubChem Structures + Assays dataset, which has a relatively small overlap with the other datasets included in the database).

The collected statistics are not necessarily representative for the quality of a given dataset, but rather provide some insight on the overall error rate and highlight the need for more work on database curation and quality assurance.

The above-mentioned 67 datasets are included in the final OpenTox database and are accessible online at https://ambit.uni-plovdiv.bg:8443/ambit2/dataset through the OpenTox REST API[90]. In particular, these datasets are available through the OpenTox Dataset API[91] and can be accessed through 3rd party web services and/or client implementing this API. They can be retrieved as complete datasets or via various kinds of search operations over chemical names, registry identifiers, structures, sub-structures (SMARTS)[92], similarity, Daylight

[90] www.opentox.org/dev/apis/api-1.1 accessed on May 27, 2011

[91] www.opentox.org/dev/apis/api-1.1/dataset accessed on May 27, 2011

[92] www.daylight.com/dayhtml/doc/theory/theory.smarts.html accessed on May 27, 2011

SMILES[93], InChI[94], etc. Of course, proper authentication and authorisation are required for accessing the data or submitting new data through POST operations.

All these functionalities are used by the OpenTox ToxPredict web application, available at toxpredict.org, thus providing a user-friendly interface on the Web to the more general scientific community to access the data.

The full raw database statistics as reported by AmbitXT are provided in Appendix A: OpenTox final database raw statistics.

2.4 Fraunhofer ITEM Repdose database

The Fraunhofer ITEM Repdose was incorporated at the end of the project as follows:

- RepDose[95] – Fraunhofer ITEM dataset was implemented in Ambit for OpenTox containing information on 583 chemicals. The structure data was imported as SDF-File. The physical/chemical parameters and data to repeated dose studies were imported from Excel files. Each entry/endpoint in the study datasets is specified with the following information:
  Study (route, duration, reliability, study NOEL in mmol, study LOEL in mmol)
  Animal (species, strain, studied sex)
  Effected Organ
  Effect (effect sex, effect LOEL in mmol)
  Study reference

  The data is divided in 6 datasets:
  - Structure data
  - Physical/chemical data
  - Inhalation studies rat: contains 3644 entries/endpoints to repeated dose inhalation studies on rat
  - Oral studies rat: contains 7182 entries/endpoints to repeated dose oral studies on rat
  - Inhalation studies mouse: contains 1195/endpoints entries to repeated dose inhalation studies on mouse
  - Oral studies mouse: contains 2226/endpoints entries to repeated dose oral studies on mouse
- An instance of AMBIT services, implementing OpenTox API was installed at the Fraunhofer ITEM server, and Repdose data imported via standard OpenTox API for uploading data and chemical structures. This makes the data available to OpenTox partners via the uniform OpenTox interface, and at the same time, ensures the hosting is controlled by the data providers (Fraunhofer ITEM). The access is restricted to OpenTox partners and an appropriate license URI is included in every dataset.

3. OpenTox Toxicological Endpoints Ontology

3.1 ECHA endpoints ontology

We have developed an OWL (Web Ontology Language)[96] ontology of toxicological endpoints[97] that corresponds to the endpoint classification of REACH guidance documents[98] and allows for unique mapping between

[93] www.daylight.com/smiles accessed on May 27, 2011

[94] www.iupac.org/inchi accessed on May 27, 2011

[95] http://www.fraunhofer-repdose.de//accessed on October 31, 2011

[96] www.w3.org/TR/owl-features accessed on May 27, 2011

endpoints from various inventories. Figure 2 provides a glimpse of the ECHA endpoints ontology summary graph, while Figure 3 illustrates its human health specific part.

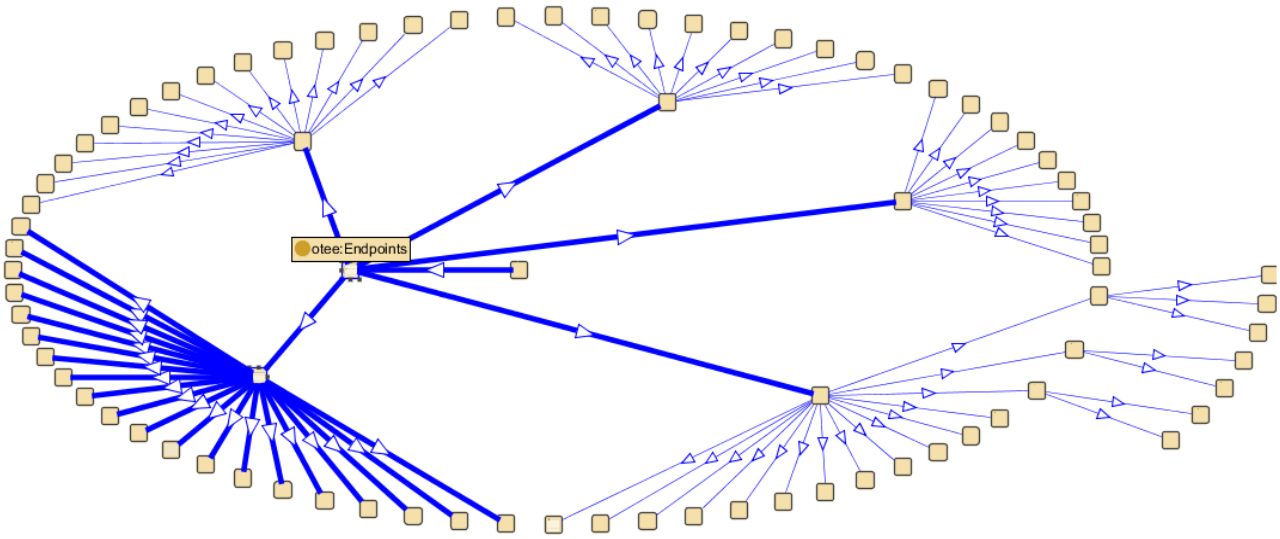

Figure 2 ECHA endpoints ontology summary graph

The quality of a (Q)SAR model crucially depends on the clarity of endpoints and experimental protocols used and the ability to communicate this information in an unambiguous way both in the model development and model application. The current common practice usually includes a textual description of the materials and methods used for acquiring experimental data as well as literature references, while the model description is a separate entity. Providing an automatic and unique way of describing and linking the endpoint information in a formal way, ready for software processing with minimal human intervention, is one of the big challenges that OpenTox's distributed web services framework tries to address. This is achieved currently by making use of a simple ontology of endpoints. We have defined an ontology based on the OWL (Web Ontology Language) for toxicological endpoints that is in line with current ECHA REACH guidance. Using this ontology, each attribute in a toxicological dataset can be associated with an entry in the ontology, therefore allowing a unique mapping between endpoints in various and heterogeneous datasets. The mapping of chemical compound properties stored in the final OpenTox database with the endpoints ontology, and the information which properties are predicted by models available via the OpenTox model service, is used to automatically recognise which endpoints have predictive models available, and ensures consistency of the used endpoint terminology across the set of distributed OpenTox services.

[97] opentox.org/data/documents/development/RDF%20files/Endpoints/view accessed on May 27, 2011

[98] guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?%20vers=20_08_08 accessed on May 27, 2011

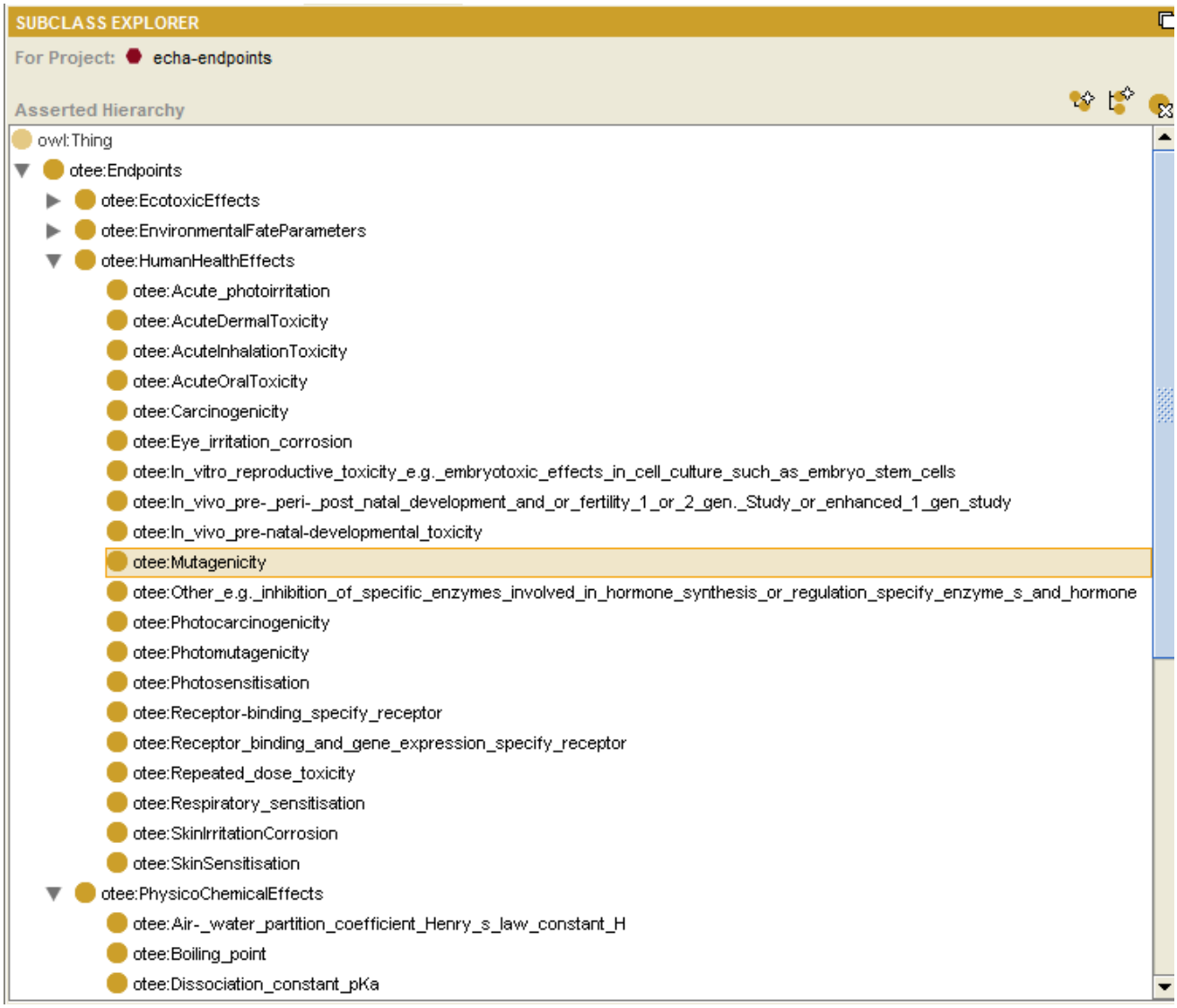

Figure 3 Human health specific part of the ECHA endpoints ontology

A full OWL dump of the Toxicological Endpoints Ontology is provided in Appendix C: OpenTox Toxicological Endpoints . The ontology has been used for mapping the relevant OpenTox database dataset fields with the ECHA endpoints. This allows for dynamic linking of models, datasets and endpoints through OpenTox API-compliant operations. The ontology service[99] stores the endpoints ontology, along with other OpenTox-relevant ontologies (opentox.owl[100], algorithm types[101] and Blue Obelisk Descriptor Ontology[102]), as well as dynamic information of available models, algorithms and features, provided by OpenTox services. The ToxPredict application[103] queries[104] the ontology service via a standard SPARQL[105] interface for all available

[99] http://apps.ideaconsult.net:8080/ontology accessed on May 27, 2011

[100] opentox.org/data/documents/development/RDF%20files/OpenToxOntology/view accessed on May 27, 2011

[101] opentox.org/data/documents/development/RDF%20files/AlgorithmTypes/view accessed on May 27, 2011

[102] opentox.org/data/documents/development/RDF%20files/BlueObeliskOntology/view accessed on May 27, 2011

[103] toxpredict.org accessed on May 27, 2011

models and retrieves associated information about endpoint modelled, algorithms used to create models and independent and target variables used in models.

3.2 OpenTox Toxicological Ontology

The construction of a formal ontology follows relatively established principles in knowledge representation. We have taken into consideration those available for biomedical ontology development, particularly the OBO Foundry principles[106] together with the ToxML (Toxicology XML standard) schema[107]. An open, public approach to ontology development supports current and future collaborations with different projects. We use the DL species of the Web Ontology Language (OWL DL) supported by the Protégé OWL editor. An overview of the OpenTox ontology is given on the public area of the OpenTox website[108] together with instructions on how to enter the OpenTox Collaborative Protégé[109] Server and contribute to existing OWL projects.

Up to now, several ontology sub-projects have been made available through the OpenTox Collaborative Protégé Server, among them: Toxicological Ontology, Organ System Ontology and the ToxML Ontology.

The OpenTox Toxicological Ontology at the moment contains five toxicity study types: carcinogenicity, *in vitro* bacterial mutagenesis, *in vivo* micronucleus, repeated dose toxicity (e.g., chronic, sub-chronic or sub-acute study types) and aquatic toxicity (see Figure 4). The purpose of this ontology is to enable the attributes of toxicological dataset entries to be associated with ontology concepts. The main OWL classes are "ToxicityStudyType", "TestSystem" (includes subclasses such as strains, species, sex, route of exposure, etc.), "TestResult" (includes subclasses such as toxicity measure, test call, mode of action, target sites, etc.). The aquatic toxicity ontology was based on the requirements of the directive of the European Union 92/69/EEC (O.J. L383 A), i.e., acute toxicity for fish (method C.1.), acute toxicity for Daphnia (C.2.), and the algal growth inhibition test (C.3.).

---

[104] opentox.org/data/documents/development/RDF%20files/JavaOnly/query-reasoning-with-jena-and-sparql accessed on May 27, 2011

[105] http://www.w3.org/TR/rdf-sparql-query/ accessed on May 27, 2011

[106] http://www.obofoundry.org/ The OBO Foundry accessed on May 27, 2011

[107] www.leadscope.com/toxml.php Leadscope ToxML Schema accessed on May 27, 211

[108] www.opentox.org/dev/ontology OpenTox ontology development page accessed on May 27, 2011

[109] http://bio-ontologies.knowledgeblog.org/187 OpenTox Predictive Toxicology Framework: toxicological ontology and semantic media wiki-based OpenToxipedia, Bio-Ontologies 2011, Vienna, Austria, July 15-16 2011

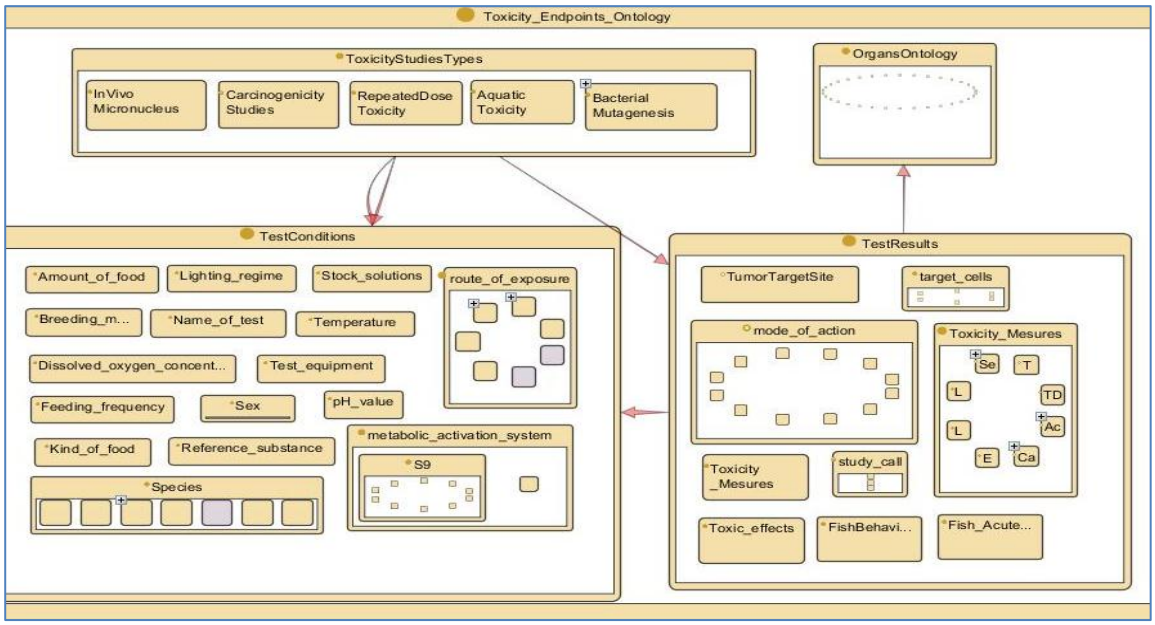

Figure 4 OpenTox toxicological ontology structure

3.3 OpenTox Organs and Effects Ontology

The "target sites" toxicological class has been extended by the Organ System Ontology, developed by the Fraunhofer Institute for Toxicology & Experimental Medicine. The Organ System Ontology is one of the most challenging ontology classes addressing targets, examinations and organs observed in *in vivo* studies such as repeated dose toxicity and carcinogenicity. The ontology includes the detailed description of organs starting from organ systems down to histological components. It was decided to use a hierarchical structure starting with the organ system (e.g. digestive system) instead of orientating the ontology on the examinations performed in guideline studies such as histopathology, necropsy, and clinical observations. Thus, the principal structure of the Organ System Ontology is as follows:

- Class Organs system – Subclass Organs system
  - Class Target organs – Subclass Target organs 1 to N
    - Class Histopathology – Subclasses if needed

At the moment the Organ System Ontology includes 13 organ systems, namely the digestive, respiratory, circulatory, endocrine, male genital, female genital, hematopoietic, integumentary, body cavities, nervous system and special sense organs, urinary, musculoskeletal, immune system, and lymphatic organs. Synonyms are included to account for differences in terminologies. The ontology focuses on the organs observed in rodents, which are frequently used for toxicity testing. Species specificity will be introduced when combining the organ ontology with the toxicological endpoint ontology.

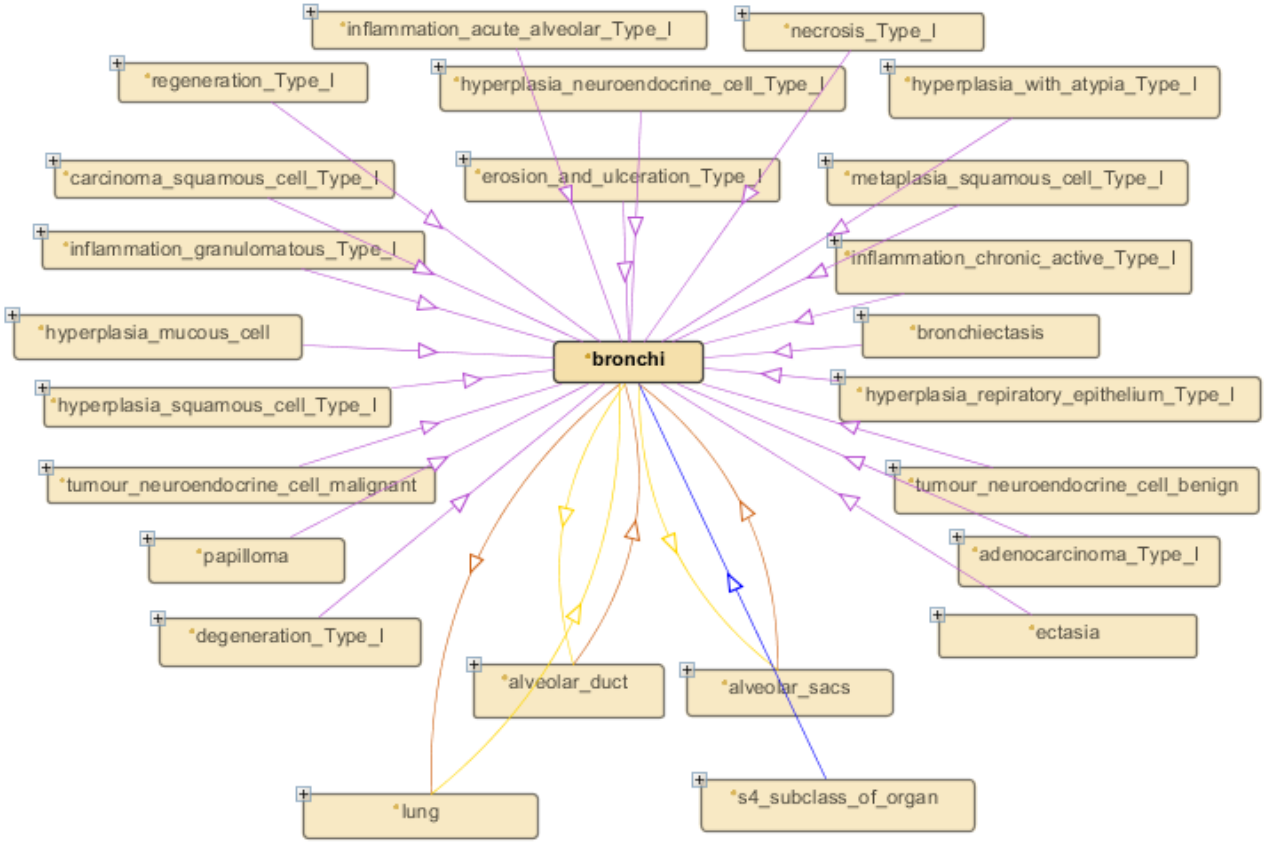

Figure 5 Example of the target site description (Organ System Ontology) for bronchi

Currently, the Toxicological Effects Ontology comprises neoplastic and non-neoplastic effects observed in repeated dose and cancer studies. Endpoint specificity of the effects will be included when combining the organ/effect root ontology with the Toxicological Endpoint Ontology. The Toxicological Effects Ontology consists of three main parts: classes of effects, linked to pathological effects, which are further linked to detailed diagnostic features as agreed in the INHAND initiative[110]. Its functionality has been initially developed for the respiratory tract. The structure of the combined organ and effects ontology is depicted in Figure 6.

[110] R. Renne , A. Brix et al. (2009), Proliferative and Nonproliferative Lesions of the Rat and Mouse Respiratory Tract, Toxicologic Pathology, 37: 5-73

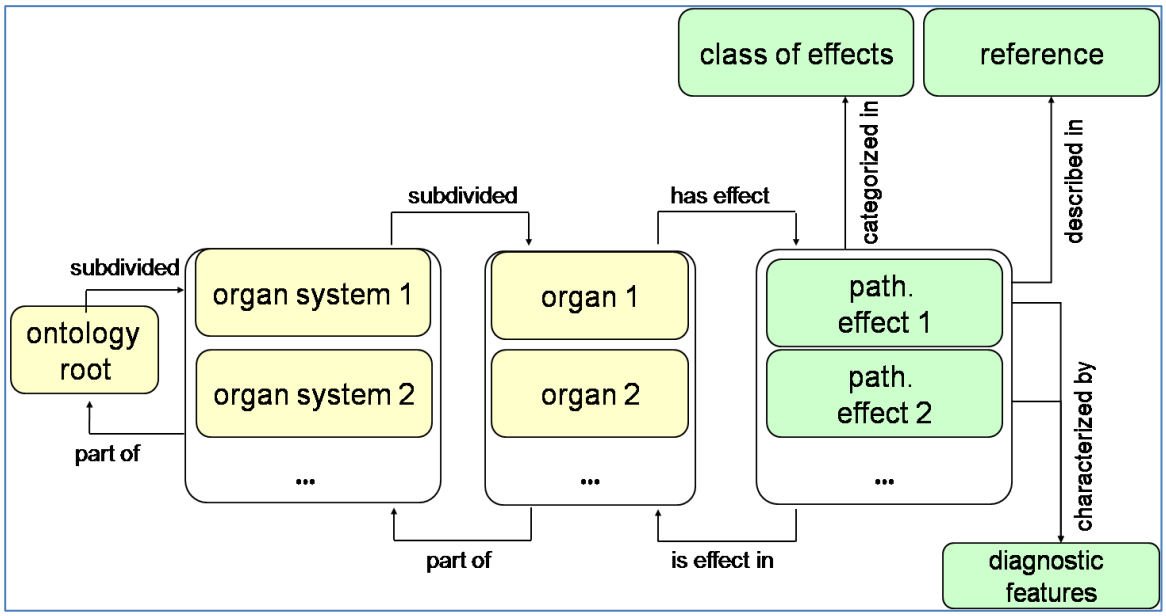

Figure 6 Overview of the structure of the combined organ (in orange) and effect (in green) ontologies

3.4 Automatic conversion of ToxML schema to an OWL ontology

ToxML is a well-recognized and very detailed schema describing toxicity study data. After some manual mappings of the ToxML fields and the databases of interest we determined that ToxML was suitable for accommodating/referencing the toxicity data. There are several issues which are considered during the automatic conversion and are explained below:

- In the XML description many fields exist with free text instead of named concepts. For example, in carcinogenicity studies data the survival rate is described as "*MORTALITY, INCREASED*" which is a plain text string not referring to any conceptual knowledge. For solving this issue in an ontology, the same values are assigned as properties of a concept which is located within a domain knowledge model. This allows referencing between known entities.
- Mixed type values (string / integer / interval) sometimes describe the same indicator in different studies. For example, in data from different chronic studies, the same indicator survival rate is presented as either of the following strings:
  '0/5 (weeks 27-30)', '10/10', '16/17', '5/5', '6/6' Representing in the first case the number of survived animals among the tested ones during some concrete period of the study and in all other cases simply the number of survived animals in relation to all tested animals. These are inconsistent data which show the same indicator value in different value types. Organizing the type values as restricted values in an ontology allows for comparison and further automatic processing which is not really feasible with plain strings.
- XML-encoded data does not use standardized concept terms for many classes, which could be easily organized in controlled vocabularies or already exist as such: target sites; mode of action; route of exposure.
- in the XML schema sometimes the same field names are used to label classes and properties. Introducing controlled vocabularies helps avoiding term overlapping and thus ambiguity.
- the XML-nested structure does not follow the natural IS-A relation used for subclassing in OWL. See the figure below, which clearly shows the discrepancies between the XML and OWL representation of the same facts.

To convert the XML schema to an OWL structure several *ad hoc* rules were implemented:

- for distinguishing classes from properties among the XML fields;

- for introducing object properties – by default in the schema all properties correspond to datatype properties in OWL because they connect an entity to a string value;
- for removal of some container classes which are not needed in an ontology (Tests, Compounds, etc.) – these are necessary in XML because they frame a set of subfields, but in OWL each test or compound is a separate object. Many of these objects can exist independently and they are all related to their originating type class;
- for renaming classes which appear with the same name in different contexts.

The resulting ontology has a flat structure representing numerous relations expressing the nested structure of the XML schema.

The IS-A relation is introduced only to a limited number of classes, for example, **ChronicStudies** *rdfs:subClassOf* **Study**

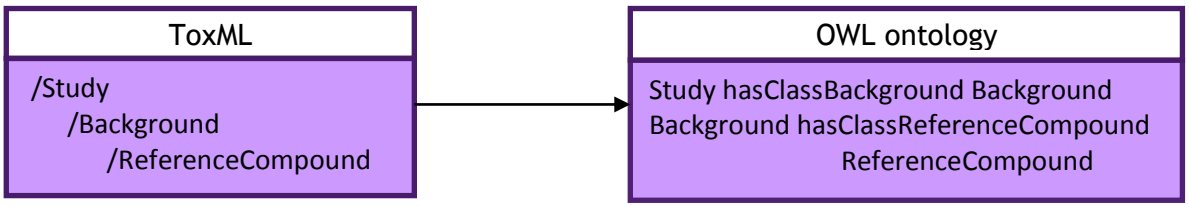

Figure 7 Introducing object type properties for each tuple of nested classes

The relations between the classes are obtained from the nested XML structures and encoded as follows: for each tuple of nested fields in the schema F_1 and F_2, two new classes in OWL are created as Class F_1 and Class F_2 and an object property *hasClass* F_2 which expresses the relation between both classes. For example, see Figure 7, which shows the corresponding representation of the nested fields Study, Background and ReferenceCoumpound on the left to OWL on the right.

Fields that have string values in XML are converted to properties in OWL. For example, the field *Name*, which has a string value type, becomes the datatype property *hasName.*

Ambiguous labels are unified: for example, the field *Results* is used in several different contexts, which is why we renamed it to *TreatmentResults* in the treatment group.

Where possible, object type properties are introduced instead of datatype ones – thus, string values are replaced by named concepts. For example, in the ToxML schema the field *Sex* is defined as a simple type of type string, which would be converted to a property in a direct conversion, but in OWL we introduced a new class *Sex*, thus creating an object property *hasSex* instead of a datatype one.

Mapping to Organ System Ontology and study type classifications can be applied in a similar manner.

4. Conclusions

The OpenTox final database is a major contribution beyond the state-of-the-art in the area of life sciences data warehousing and provides a solid basis for the OpenTox framework of RESTful web services. It enables improved storage, exchange, aggregation, quality labelling, curation and integrated use of high-quality life sciences information, and allows for consistent and scientifically sound mathematical and computer modelling, including modelling of structure-activity relationships for REACH-relevant endpoints.

The ToxPredict web application, which estimates the chemical hazard of chemical structures and is available at toxpredict.org, illustrates one generic use case for the OpenTox final database. Users can either search the OpenTox final database, which currently includes quality-labelled data for 646,042 chemicals, grouped in 67 datasets, or upload their own chemical structure data and eventually choose to allow third parties to work on it

in a collaborative way. ToxPredict enables transparent and sound scientific modelling of chemical properties, in compliance with the best current practices for data management and records keeping, implemented in the final OpenTox database.

Providing an automatic and unique approach of describing and linking the endpoint information in a formal way, ready for software processing with minimal human intervention, is one of the big challenges that OpenTox's distributed web services framework tries to address. Using the ontology, each attribute in a toxicological dataset can be associated with an entry in the ontology, therefore allowing a unique mapping between endpoints in various and heterogeneous datasets. The mapping of chemical compound properties stored in the OpenTox database with the endpoints ontology, and the information which properties are predicted by models available via the OpenTox model service, are used to automatically recognise which endpoints have predictive models available and ensure consistency of the used endpoint terminology across the set of distributed OpenTox services.

Another important aspect of the final OpenTox database is that it could be either hosted on a single computer (even a laptop or netbook would do fine) or could be transparently distributed on multiple servers in various physical locations, in particular for better reliability, resilience, performance and scalability. Moreover, it could be deployed behind firewalls, on Intranets (or even offline), when very tight security policies have to be met. Third parties, willing to deploy the final OpenTox database in-house, could select to install a relevant subset of datasets, tailored to their specific needs.

The OpenTox approach to data resource management and integration addresses the replacement alternatives challenge in a multi-domain friendly way, which is essential for data and model sharing, repeatability and validation of prediction results – really a must in 21st century toxicological science.

Appendix A: OpenTox final database raw statistics

The final OpenTox database raw statistics as reported by AmbitXT are reported below:

```
Number of datasets     67


Dataset Number of compounds
ECHA list of pre-registered substances (20090327)    143835
ChemIDplus (20110503)  80468
Chemical Identifier Resolver (20110212)      72985
ChemDraw (20110505)    22519
CPDBAS  1547
DBPCAN  209
EPAFHM  617
FDAMDD  1216
HPVCSI  3548
HPVISD  1006
IRISTR  544
KIERBL  278
NCTRER  232
NTPBSI  2330
NTPHTS  1408
ISSCAN  1150
ISSMIC  151
ISSSTY  223
TOXCST  320
TXCST2  960
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) 176

Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data
(Dermatitis Vol 16 No 4 2005) 209
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol
21 No 1 2010)  108
Bioconcentration factor (BCF) Gold Standard Database 1130
Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way 185
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity    6512
Bursi AMES Toxicity Dataset    4337
EPI_AOP 818
EPI_BCF 685
EPI_BioHC      175
EPI_Biowin     1263
EPI_Boil_Pt    5890
EPI_Henry      1829
EPI_KM  631
EPI_KOA 308
EPI_Kowwin     15809
EPI_Melt_Pt    10051
EPI_PCKOC      788
EPI_VP  3037
EPI_WaterFrag  5764
EPI_Wskowwin   2348
TOXCST_ACEA    320
TOXCST_Attagene        320
TOXCST_BioSeek 320
TOXCST_Cellumen        320
TOXCST_CellzDirect     320
TOXCST_Gentronix       320
TOXCST_NCGC    320
TOXCST_Novascreen      320
TOXCST_Solidus 320
TOXCST_ToxRefDB        320
ECBPRS (20090917)      80410
NAME2STRUCTURE 70646
PubChem Structures + Assays (20090924)       473965
Leadscope_carc_level_2 2988
Leadscope_ccris_genetox        8001
Leadscope_cder_chronic 121
Leadscope_cder_genetox 336
Leadscope_cder_repro_dev       58
Leadscope_cfsan_acute  1070
Leadscope_cfsan_chronic        655
```

Leadscope_cfsan_genetox	696
Leadscope_cfsan_repro_dev	312
Leadscope_fda_marketed_drugs	6637
Leadscope_genetox_level_2	10155
Leadscope_ntp_genetox	2128
Pharmatrope_AERS_hepatobiliary_system	1274

Dataset	Number of empty structures
ECHA list of pre-registered substances (20090327)	143835
CPDBAS	39
HPVCSI	1068
HPVISD	302
IRISTR	5
NTPBSI	151
NTPHTS	14
ISSCAN	12
ISSMIC	6
ISSSTY	1
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	1

Bioconcentration factor (BCF) Gold Standard Database	1130
EPI_Boil_Pt	1
EPI_Melt_Pt	5
EPI_PCKOC	2
EPI_VP	1
EPI_WaterFrag	3
EPI_Wskowwin	3
NAME2STRUCTURE	16
PubChem Structures + Assays (20090924)	45
Leadscope_carc_level_2	141
Leadscope_ccris_genetox	559
Leadscope_cder_chronic	4
Leadscope_cder_genetox	242
Leadscope_cfsan_acute	100
Leadscope_cfsan_chronic	106
Leadscope_cfsan_genetox	130
Leadscope_cfsan_repro_dev	47
Leadscope_fda_marketed_drugs	321
Leadscope_genetox_level_2	748
Leadscope_ntp_genetox	71

Quality labels summary Entire database

Label	Details	Number of chemicals
Consensus	10	615
Consensus	11	452
Consensus	12	329
Consensus	13	310
Consensus	14	206
Consensus	15	177
Consensus	16	128
Consensus	17	157
Consensus	18	115
Consensus	19	93
Consensus	2	12272
Consensus	20	80
Consensus	21	72
Consensus	22	79
Consensus	23	51
Consensus	24	65
Consensus	25	49
Consensus	26	46
Consensus	27	49
Consensus	28	24
Consensus	29	28
Consensus	3	10311
Consensus	30	20
Consensus	31	23
Consensus	32	16
Consensus	33	14
Consensus	34	10

Consensus	35	7
Consensus	36	10
Consensus	37	1
Consensus	38	2
Consensus	39	3
Consensus	4	26390
Consensus	40	1
Consensus	42	1
Consensus	46	1
Consensus	5	10908
Consensus	6	3552
Consensus	7	2080
Consensus	8	1208
Consensus	9	780
Majority	10:2	11
Majority	10:3	11
Majority	10:4	6
Majority	10:5	5
Majority	10:6	2
Majority	10:7	1
Majority	10:8	1
Majority	11:14	1
Majority	11:2	8
Majority	11:3	15
Majority	11:4	13
Majority	11:5	5
Majority	11:6	2
Majority	11:7	2
Majority	12:2	4
Majority	12:2:4	1
Majority	12:3	13
Majority	12:4	9
Majority	12:5	3
Majority	12:6	2
Majority	13:2	4
Majority	13:3	9
Majority	13:4	16
Majority	13:5	3
Majority	13:6	3
Majority	14:3	7
Majority	14:4	7
Majority	14:5	9
Majority	14:6	4
Majority	14:8	1
Majority	15:2	1
Majority	15:3	4
Majority	15:4	3
Majority	15:5	1
Majority	15:6	2
Majority	16:3	6
Majority	16:4	4
Majority	16:5	2
Majority	17:2	3
Majority	17:3	4
Majority	17:4	3
Majority	17:5	2
Majority	17:6	1
Majority	17:8	1
Majority	18:3	2
Majority	18:4	1
Majority	18:5	3
Majority	18:6	1
Majority	19:2	1
Majority	19:3	2
Majority	19:4	4
Majority	19:5	4
Majority	1:10	38
Majority	1:10:2	1
Majority	1:10:3	1
Majority	1:11	31
Majority	1:11:2	2
Majority	1:11:4	1
Majority	1:12	26
Majority	1:12:4	2

Majority	1:13	19
Majority	1:13:3	1
Majority	1:14	11
Majority	1:14:6	1
Majority	1:15	8
Majority	1:15:2	1
Majority	1:16	8
Majority	1:16:5	1
Majority	1:17	3
Majority	1:17:9	1
Majority	1:18	8
Majority	1:18:6	1
Majority	1:19	2
Majority	1:1:10	2
Majority	1:1:11	3
Majority	1:1:11:6	1
Majority	1:1:12	2
Majority	1:1:13	1
Majority	1:1:13:6	1
Majority	1:1:16	1
Majority	1:1:16:7	1
Majority	1:1:1:16	1
Majority	1:1:1:17	1
Majority	1:1:1:1:3	2
Majority	1:1:1:2	28
Majority	1:1:1:2:4	1
Majority	1:1:1:3	13
Majority	1:1:1:4	1
Majority	1:1:1:5	1
Majority	1:1:1:6	1
Majority	1:1:1:7	1
Majority	1:1:2	1178
Majority	1:1:24	1
Majority	1:1:2:3	3
Majority	1:1:2:4	1
Majority	1:1:3	274
Majority	1:1:3:4	1
Majority	1:1:3:7	1
Majority	1:1:3:8	1
Majority	1:1:4	51
Majority	1:1:4:5	1
Majority	1:1:4:8	1
Majority	1:1:5	34
Majority	1:1:6	8
Majority	1:1:7	10
Majority	1:1:8	7
Majority	1:1:9	2
Majority	1:2	3690
Majority	1:20	1
Majority	1:21	3
Majority	1:22	1
Majority	1:23	1
Majority	1:24	1
Majority	1:27	2
Majority	1:29	1
Majority	1:2:2:5	2
Majority	1:2:3	35
Majority	1:2:3:3:4	1
Majority	1:2:4	19
Majority	1:2:5	13
Majority	1:2:6	12
Majority	1:2:7	4
Majority	1:2:8	2
Majority	1:3	6007
Majority	1:30	2
Majority	1:3:4	17
Majority	1:3:5	8
Majority	1:3:6	3
Majority	1:3:7	8
Majority	1:3:8	5
Majority	1:3:9	1
Majority	1:4	2003
Majority	1:4:4:8	1
Majority	1:4:5	2

Majority	1:4:6	2
Majority	1:4:7	2
Majority	1:4:8	3
Majority	1:4:9	2
Majority	1:5	506
Majority	1:5:7	3
Majority	1:5:8	1
Majority	1:5:9	2
Majority	1:6	250
Majority	1:6:7	2
Majority	1:7	164
Majority	1:8	73
Majority	1:8:9	1
Majority	1:9	67
Majority	20:3	1
Majority	20:4	3
Majority	20:5	2
Majority	20:6	1
Majority	21:4	3
Majority	21:5	6
Majority	22:4	1
Majority	22:5	4
Majority	22:8	1
Majority	23:4	1
Majority	23:5	4
Majority	23:6	1
Majority	24:3	1
Majority	24:4	1
Majority	24:5	1
Majority	25:3	1
Majority	25:4	2
Majority	25:5	1
Majority	26:3	1
Majority	26:4	2
Majority	26:5	2
Majority	27:6	1
Majority	2:20	2
Majority	2:2:3	1
Majority	2:2:4	2
Majority	2:2:5	1
Majority	2:2:6	1
Majority	2:2:7	2
Majority	2:2:8	1
Majority	2:3	516
Majority	2:3:4	2
Majority	2:3:7	2
Majority	2:3:9	1
Majority	2:4	157
Majority	2:4:5	1
Majority	2:4:5:9	1
Majority	2:4:6	1
Majority	2:4:7	1
Majority	2:4:8	1
Majority	2:4:9	2
Majority	2:5	58
Majority	2:5:8	1
Majority	2:6	54
Majority	2:7	30
Majority	2:8	25
Majority	2:9	16
Majority	3:3:6	1
Majority	3:4	102
Majority	3:4:7	1
Majority	3:5	59
Majority	3:5:7	1
Majority	3:6	39
Majority	3:7	27
Majority	3:8	31
Majority	3:9	21
Majority	4:4:5	1
Majority	4:5	29
Majority	4:5:6	1
Majority	4:6	20
Majority	4:7	14

Majority	4:8	15
Majority	4:9	16
Majority	5:6	11
Majority	5:7	9
Majority	5:8	3
Majority	5:9	7
Majority	6:7	6
Majority	6:8	12
Majority	6:9	3
Majority	7:8	2
Unconfirmed	1	498953
Ambiguous	1:1	2220
Ambiguous	1:1:1	485
Ambiguous	1:1:1:1	79
Ambiguous	1:1:1:1:1	10
Ambiguous	1:1:1:2:2	1
Ambiguous	1:1:2:2	8
Ambiguous	1:1:3:3	2
Ambiguous	1:2:2	93
Ambiguous	1:2:3:3	1
Ambiguous	1:3:3	13
Ambiguous	1:4:4	5
Ambiguous	1:6:6	1
Ambiguous	2:2	1227
Ambiguous	2:2:2	1
Ambiguous	2:2:2:2	1
Ambiguous	2:4:4	1
Ambiguous	3:3	82
Ambiguous	4:4	26
Ambiguous	5:5	6
Ambiguous	6:6	5
Ambiguous	7:7	2

Dataset	Mode	label	Number of compounds
ChemIDplus (20110503)	comparison	OK	59620
ChemIDplus (20110503)	comparison	ProbablyOK	12694
ChemIDplus (20110503)	comparison	Unknown	5497
ChemIDplus (20110503)	comparison	ProbablyERROR	2657
Chemical Identifier Resolver (20110212)	comparison	OK	55470
Chemical Identifier Resolver (20110212)	comparison	ProbablyOK	9719
Chemical Identifier Resolver (20110212)	comparison	Unknown	3164
Chemical Identifier Resolver (20110212)	comparison	ProbablyERROR	4631
ChemDraw (20110505)	comparison	OK	18147
ChemDraw (20110505)	comparison	ProbablyOK	3542
ChemDraw (20110505)	comparison	Unknown	263
ChemDraw (20110505)	comparison	ProbablyERROR	567
CPDBAS	comparison	OK	1011
CPDBAS	comparison	ProbablyOK	372
CPDBAS	comparison	Unknown	78
CPDBAS	comparison	ProbablyERROR	47
DBPCAN	comparison	OK	150
DBPCAN	comparison	ProbablyOK	5
DBPCAN	comparison	Unknown	54
EPAFHM	comparison	OK	510
EPAFHM	comparison	ProbablyOK	78
EPAFHM	comparison	Unknown	28
EPAFHM	comparison	ProbablyERROR	1
FDAMDD	comparison	OK	759
FDAMDD	comparison	ProbablyOK	291
FDAMDD	comparison	Unknown	148
FDAMDD	comparison	ProbablyERROR	18
HPVCSI	comparison	OK	1983
HPVCSI	comparison	ProbablyOK	342
HPVCSI	comparison	Unknown	145
HPVCSI	comparison	ProbablyERROR	10
HPVISD	comparison	OK	603
HPVISD	comparison	ProbablyOK	98
HPVISD	comparison	Unknown	3
IRISTR	comparison	OK	414
IRISTR	comparison	ProbablyOK	99
IRISTR	comparison	Unknown	23
IRISTR	comparison	ProbablyERROR	3

KIERBL comparison OK 224
KIERBL comparison ProbablyOK 48
KIERBL comparison Unknown 5
KIERBL comparison ProbablyERROR 1
NCTRER comparison OK 180
NCTRER comparison ProbablyOK 17
NCTRER comparison Unknown 35
NTPBSI comparison OK 1583
NTPBSI comparison ProbablyOK 520
NTPBSI comparison Unknown 48
NTPBSI comparison ProbablyERROR 28
NTPHTS comparison OK 1062
NTPHTS comparison ProbablyOK 322
NTPHTS comparison Unknown 1
NTPHTS comparison ProbablyERROR 9
ISSCAN comparison OK 714
ISSCAN comparison ProbablyOK 265
ISSCAN comparison Unknown 134
ISSCAN comparison ProbablyERROR 25
ISSMIC comparison OK 105
ISSMIC comparison ProbablyOK 32
ISSMIC comparison Unknown 4
ISSMIC comparison ProbablyERROR 4
ISSSTY comparison OK 157
ISSSTY comparison ProbablyOK 55
ISSSTY comparison Unknown 5
ISSSTY comparison ProbablyERROR 5
TOXCST comparison OK 253
TOXCST comparison ProbablyOK 60
TOXCST comparison Unknown 1
TOXCST comparison ProbablyERROR 3
TXCST2 comparison OK 677
TXCST2 comparison ProbablyOK 156
TXCST2 comparison Unknown 121
TXCST2 comparison ProbablyERROR 6
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison OK 163
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison ProbablyOK 9
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison Unknown 2
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison ProbablyERROR 1
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) comparison OK 125
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) comparison ProbablyOK 16
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) comparison Unknown 67
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) comparison ProbablyERROR 1
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010) comparison OK 66
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010) comparison ProbablyOK 12
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010) comparison Unknown 29
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010) comparison ProbablyERROR 1
Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way comparison OK 50
Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way comparison ProbablyOK 7
Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way comparison Unknown 128
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity comparison OK 3687
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity comparison ProbablyOK 428

Benchmark Data Set for In Silico Prediction of Ames Mutagenicity comparison Unknown 2204
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity comparison ProbablyERROR 193

Bursi AMES Toxicity Dataset comparison OK 2311
Bursi AMES Toxicity Dataset comparison ProbablyOK 587
Bursi AMES Toxicity Dataset comparison Unknown 1071

```
Bursi AMES Toxicity Dataset   comparison     ProbablyERROR  368
EPI_AOP comparison     OK      713
EPI_AOP comparison     ProbablyOK     39
EPI_AOP comparison     Unknown 61
EPI_AOP comparison     ProbablyERROR  5
EPI_BCF comparison     OK      547
EPI_BCF comparison     ProbablyOK     107
EPI_BCF comparison     Unknown 22
EPI_BCF comparison     ProbablyERROR  9
EPI_BioHC      comparison     OK      159
EPI_BioHC      comparison     Unknown 16
EPI_Biowin     comparison     OK      1045
EPI_Biowin     comparison     ProbablyOK     157
EPI_Biowin     comparison     Unknown 53
EPI_Biowin     comparison     ProbablyERROR  8
EPI_Boil_Pt    comparison     OK      5124
EPI_Boil_Pt    comparison     ProbablyOK     299
EPI_Boil_Pt    comparison     Unknown 436
EPI_Boil_Pt    comparison     ProbablyERROR  30
EPI_Henry      comparison     OK      1605
EPI_Henry      comparison     ProbablyOK     203
EPI_Henry      comparison     Unknown 15
EPI_Henry      comparison     ProbablyERROR  6
EPI_KM  comparison     OK      511
EPI_KM  comparison     ProbablyOK     67
EPI_KM  comparison     Unknown 53
EPI_KOA comparison     OK      244
EPI_KOA comparison     ProbablyOK     15
EPI_KOA comparison     Unknown 49
EPI_Kowwin     comparison     OK      5730
EPI_Kowwin     comparison     ProbablyOK     956
EPI_Kowwin     comparison     Unknown 8991
EPI_Kowwin     comparison     ProbablyERROR  132
EPI_Melt_Pt    comparison     OK      7864
EPI_Melt_Pt    comparison     ProbablyOK     1167
EPI_Melt_Pt    comparison     Unknown 812
EPI_Melt_Pt    comparison     ProbablyERROR  203
EPI_PCKOC      comparison     OK      616
EPI_PCKOC      comparison     ProbablyOK     124
EPI_PCKOC      comparison     Unknown 36
EPI_PCKOC      comparison     ProbablyERROR  10
EPI_VP  comparison     OK      2697
EPI_VP  comparison     ProbablyOK     286
EPI_VP  comparison     Unknown 42
EPI_VP  comparison     ProbablyERROR  11
EPI_WaterFrag  comparison     OK      3980
EPI_WaterFrag  comparison     ProbablyOK     588
EPI_WaterFrag  comparison     Unknown 1119
EPI_WaterFrag  comparison     ProbablyERROR  74
EPI_Wskowwin   comparison     OK      2001
EPI_Wskowwin   comparison     ProbablyOK     331
EPI_Wskowwin   comparison     Unknown 6
EPI_Wskowwin   comparison     ProbablyERROR  7
TOXCST_ACEA    comparison     OK      253
TOXCST_ACEA    comparison     ProbablyOK     60
TOXCST_ACEA    comparison     Unknown 1
TOXCST_ACEA    comparison     ProbablyERROR  3
TOXCST_Attagene        comparison     OK      253
TOXCST_Attagene        comparison     ProbablyOK     60
TOXCST_Attagene        comparison     Unknown 1
TOXCST_Attagene        comparison     ProbablyERROR  3
TOXCST_BioSeek comparison     OK      253
TOXCST_BioSeek comparison     ProbablyOK     60
TOXCST_BioSeek comparison     Unknown 1
TOXCST_BioSeek comparison     ProbablyERROR  3
TOXCST_Cellumen        comparison     OK      253
TOXCST_Cellumen        comparison     ProbablyOK     60
TOXCST_Cellumen        comparison     Unknown 1
TOXCST_Cellumen        comparison     ProbablyERROR  3
TOXCST_CellzDirect     comparison     OK      253
TOXCST_CellzDirect     comparison     ProbablyOK     60
TOXCST_CellzDirect     comparison     Unknown 1
TOXCST_CellzDirect     comparison     ProbablyERROR  3
TOXCST_Gentronix       comparison     OK      253
```

TOXCST_Gentronix	comparison	ProbablyOK	60	
TOXCST_Gentronix	comparison	Unknown	1	
TOXCST_Gentronix	comparison	ProbablyERROR	3	
TOXCST_NCGC	comparison	OK	253	
TOXCST_NCGC	comparison	ProbablyOK	60	
TOXCST_NCGC	comparison	Unknown	1	
TOXCST_NCGC	comparison	ProbablyERROR	3	
TOXCST_Novascreen	comparison	OK	253	
TOXCST_Novascreen	comparison	ProbablyOK	60	
TOXCST_Novascreen	comparison	Unknown	1	
TOXCST_Novascreen	comparison	ProbablyERROR	3	
TOXCST_Solidus	comparison	OK	253	
TOXCST_Solidus	comparison	ProbablyOK	60	
TOXCST_Solidus	comparison	Unknown	1	
TOXCST_Solidus	comparison	ProbablyERROR	3	
TOXCST_ToxRefDB	comparison	OK	253	
TOXCST_ToxRefDB	comparison	ProbablyOK	60	
TOXCST_ToxRefDB	comparison	Unknown	1	
TOXCST_ToxRefDB	comparison	ProbablyERROR	3	
ECBPRS (20090917)	comparison	OK	56079	
ECBPRS (20090917)	comparison	ProbablyOK	12346	
ECBPRS (20090917)	comparison	Unknown	9503	
ECBPRS (20090917)	comparison	ProbablyERROR	2482	
NAME2STRUCTURE	comparison	OK	49251	
NAME2STRUCTURE	comparison	ProbablyOK	6116	
NAME2STRUCTURE	comparison	Unknown	8635	
NAME2STRUCTURE	comparison	ProbablyERROR	6628	
PubChem Structures + Assays (20090924)	comparison	OK	7920	
PubChem Structures + Assays (20090924)	comparison	ProbablyOK	1424	
PubChem Structures + Assays (20090924)	comparison	Unknown	464355	
PubChem Structures + Assays (20090924)	comparison	ProbablyERROR	221	
Leadscope_carc_level_2	comparison	OK	1899	
Leadscope_carc_level_2	comparison	ProbablyOK	285	
Leadscope_carc_level_2	comparison	Unknown	312	
Leadscope_carc_level_2	comparison	ProbablyERROR	351	
Leadscope_ccris_genetox	comparison	OK	5743	
Leadscope_ccris_genetox	comparison	ProbablyOK	875	
Leadscope_ccris_genetox	comparison	Unknown	196	
Leadscope_ccris_genetox	comparison	ProbablyERROR	628	
Leadscope_cder_chronic	comparison	OK	83	
Leadscope_cder_chronic	comparison	ProbablyOK	20	
Leadscope_cder_chronic	comparison	Unknown	5	
Leadscope_cder_chronic	comparison	ProbablyERROR	9	
Leadscope_cder_genetox	comparison	OK	75	
Leadscope_cder_genetox	comparison	ProbablyOK	14	
Leadscope_cder_genetox	comparison	Unknown	1	
Leadscope_cder_genetox	comparison	ProbablyERROR	4	
Leadscope_cder_repro_dev	comparison	OK	42	
Leadscope_cder_repro_dev	comparison	ProbablyOK	13	
Leadscope_cder_repro_dev	comparison	ProbablyERROR	3	
Leadscope_cfsan_acute	comparison	OK	806	
Leadscope_cfsan_acute	comparison	ProbablyOK	58	
Leadscope_cfsan_acute	comparison	Unknown	94	
Leadscope_cfsan_acute	comparison	ProbablyERROR	12	
Leadscope_cfsan_chronic	comparison	OK	458	
Leadscope_cfsan_chronic	comparison	ProbablyOK	54	
Leadscope_cfsan_chronic	comparison	Unknown	26	
Leadscope_cfsan_chronic	comparison	ProbablyERROR	11	
Leadscope_cfsan_genetox	comparison	OK	486	
Leadscope_cfsan_genetox	comparison	ProbablyOK	54	
Leadscope_cfsan_genetox	comparison	Unknown	15	
Leadscope_cfsan_genetox	comparison	ProbablyERROR	11	
Leadscope_cfsan_repro_dev	comparison	OK	211	
Leadscope_cfsan_repro_dev	comparison	ProbablyOK	38	
Leadscope_cfsan_repro_dev	comparison	Unknown	5	
Leadscope_cfsan_repro_dev	comparison	ProbablyERROR	11	
Leadscope_fda_marketed_drugs	comparison	OK	2579	
Leadscope_fda_marketed_drugs	comparison	ProbablyOK	465	
Leadscope_fda_marketed_drugs	comparison	Unknown	2595	
Leadscope_fda_marketed_drugs	comparison	ProbablyERROR	677	
Leadscope_genetox_level_2	comparison	OK	6970	
Leadscope_genetox_level_2	comparison	ProbablyOK	1033	
Leadscope_genetox_level_2	comparison	Unknown	646	
Leadscope_genetox_level_2	comparison	ProbablyERROR	758	

Leadscope_ntp_genetox	comparison	OK	1580	
Leadscope_ntp_genetox	comparison	ProbablyOK	140	
Leadscope_ntp_genetox	comparison	Unknown	8	
Leadscope_ntp_genetox	comparison	ProbablyERROR	329	
Pharmatrope_AERS_hepatobiliary_system	comparison	OK	685	
Pharmatrope_AERS_hepatobiliary_system	comparison	ProbablyOK	98	
Pharmatrope_AERS_hepatobiliary_system	comparison	Unknown	320	
Pharmatrope_AERS_hepatobiliary_system	comparison	ProbablyERROR	171	

Datasets	Mode	label	Number of compounds		
ECHA list of pre-registered substances (20090327)	CAS numbers	OK	118284		
ECHA list of pre-registered substances (20090327)	CAS numbers	ERROR	2		
ECHA list of pre-registered substances (20090327)	EINECS numbers	OK	143654		
ECHA list of pre-registered substances (20090327)	EINECS numbers	ERROR	181		
ChemIDplus (20110503)	CAS numbers	OK	80468		
Chemical Identifier Resolver (20110212)	CAS numbers	OK	72985		
ChemDraw (20110505)	EINECS numbers	OK	22518		
CPDBAS	CAS numbers	OK	1528		
DBPCAN	CAS numbers	OK	179		
EPAFHM	CAS numbers	OK	617		
FDAMDD	CAS numbers	OK	1216		
HPVCSI	CAS numbers	OK	3548		
HPVISD	CAS numbers	OK	1006		
IRISTR	CAS numbers	OK	536		
KIERBL	CAS numbers	OK	278		
NCTRER	CAS numbers	OK	227		
NTPBSI	CAS numbers	OK	2218		
NTPHTS	CAS numbers	OK	1407		
ISSCAN	CAS numbers	OK	1139		
ISSMIC	CAS numbers	OK	150		
TOXCST	CAS numbers	OK	320		
TXCST2	CAS numbers	OK	927		
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	CAS numbers	OK	172		
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	CAS numbers	ERROR	4		
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005)	CAS numbers	OK	203		
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010)	CAS numbers	OK	99		
Bioconcentration factor (BCF) Gold Standard Database	CAS numbers	OK	1122		
Bioconcentration factor (BCF) Gold Standard Database	CAS numbers	ERROR	7		
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity	CAS numbers	OK	6197		
Bursi AMES Toxicity Dataset	CAS numbers	OK	4337		
EPI_AOP	CAS numbers	OK	810		
EPI_BCF	CAS numbers	OK	675		
EPI_BCF	CAS numbers	ERROR	1		
EPI_BioHC	CAS numbers	OK	170		
EPI_Biowin	CAS numbers	OK	1215		
EPI_Biowin	CAS numbers	ERROR	1		
EPI_Boil_Pt	CAS numbers	OK	5859		
EPI_Boil_Pt	CAS numbers	ERROR	1		
EPI_Henry	CAS numbers	OK	1821		
EPI_KM	CAS numbers	OK	602		
EPI_KOA	CAS numbers	OK	259		
EPI_Kowwin	CAS numbers	OK	12162		
EPI_Melt_Pt	CAS numbers	OK	10028		
EPI_Melt_Pt	CAS numbers	ERROR	1		
EPI_PCKOC	CAS numbers	OK	770		
EPI_VP	CAS numbers	OK	3031		
EPI_WaterFrag	CAS numbers	OK	5753		
EPI_Wskowwin	CAS numbers	OK	2346		
EPI_Wskowwin	CAS numbers	ERROR	1		
TOXCST_ACEA	CAS numbers	OK	320		
TOXCST_Attagene	CAS numbers	OK	320		
TOXCST_BioSeek	CAS numbers	OK	320		
TOXCST_Cellumen	CAS numbers	OK	320		
TOXCST_CellzDirect	CAS numbers	OK	320		
TOXCST_Gentronix	CAS numbers	OK	320		
TOXCST_NCGC	CAS numbers	OK	320		
TOXCST_Novascreen	CAS numbers	OK	320		
TOXCST_Solidus	CAS numbers	OK	320		

TOXCST_ToxRefDB	CAS numbers	OK	320
ECBPRS (20090917)	CAS numbers	OK	80410
ECBPRS (20090917)	EINECS numbers	OK	70749
PubChem Structures + Assays (20090924)	CAS numbers	OK	1275
Leadscope_carc_level_2	CAS numbers	OK	600
Leadscope_ccris_genetox	CAS numbers	OK	1410
Leadscope_cder_chronic	CAS numbers	OK	2
Leadscope_cder_genetox	CAS numbers	OK	12
Leadscope_cder_genetox	CAS numbers	ERROR	1
Leadscope_cder_repro_dev	CAS numbers	OK	2
Leadscope_cfsan_acute	CAS numbers	OK	690
Leadscope_cfsan_chronic	CAS numbers	OK	307
Leadscope_cfsan_genetox	CAS numbers	OK	247
Leadscope_cfsan_repro_dev	CAS numbers	OK	135
Leadscope_fda_marketed_drugs	CAS numbers	OK	451
Leadscope_genetox_level_2	CAS numbers	OK	1723
Leadscope_ntp_genetox	CAS numbers	OK	956

Dataset	Mode	label	Number of compounds		
ECHA list of pre-registered substances (20090327)	EINECS numbers,CAS numbers	OK	143654		
ECHA list of pre-registered substances (20090327)	EINECS numbers,CAS numbers	ERROR	183		
ChemIDplus (20110503)	CAS numbers	OK	80468		
Chemical Identifier Resolver (20110212)	CAS numbers	OK	72985		
ChemDraw (20110505)	EINECS numbers	OK	22518		
CPDBAS	CAS numbers	OK	1528		
DBPCAN	CAS numbers	OK	179		
EPAFHM	CAS numbers	OK	617		
FDAMDD	CAS numbers	OK	1216		
HPVCSI	CAS numbers	OK	3548		
HPVISD	CAS numbers	OK	1006		
IRISTR	CAS numbers	OK	536		
KIERBL	CAS numbers	OK	278		
NCTRER	CAS numbers	OK	227		
NTPBSI	CAS numbers	OK	2218		
NTPHTS	CAS numbers	OK	1407		
ISSCAN	CAS numbers	OK	1139		
ISSMIC	CAS numbers	OK	150		
TOXCST	CAS numbers	OK	320		
TXCST2	CAS numbers	OK	927		
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	CAS numbers	OK	172		
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	CAS numbers	ERROR	4		
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005)	CAS numbers	OK	203		
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010)	CAS numbers	OK	99		
Bioconcentration factor (BCF) Gold Standard Database	CAS numbers	OK	1122		
Bioconcentration factor (BCF) Gold Standard Database	CAS numbers	ERROR	7		
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity	CAS numbers	OK	6197		
Bursi AMES Toxicity Dataset	CAS numbers	OK	4337		
EPI_AOP	CAS numbers	OK	810		
EPI_BCF	CAS numbers	OK	675		
EPI_BCF	CAS numbers	ERROR	1		
EPI_BioHC	CAS numbers	OK	170		
EPI_Biowin	CAS numbers	OK	1215		
EPI_Biowin	CAS numbers	ERROR	1		
EPI_Boil_Pt	CAS numbers	OK	5859		
EPI_Boil_Pt	CAS numbers	ERROR	1		
EPI_Henry	CAS numbers	OK	1821		
EPI_KM	CAS numbers	OK	602		
EPI_KOA	CAS numbers	OK	259		
EPI_Kowwin	CAS numbers	OK	12162		
EPI_Melt_Pt	CAS numbers	OK	10028		
EPI_Melt_Pt	CAS numbers	ERROR	1		
EPI_PCKOC	CAS numbers	OK	770		
EPI_VP	CAS numbers	OK	3031		
EPI_WaterFrag	CAS numbers	OK	5753		
EPI_Wskowwin	CAS numbers	OK	2346		
EPI_Wskowwin	CAS numbers	ERROR	1		
TOXCST_ACEA	CAS numbers	OK	320		
TOXCST_Attagene	CAS numbers	OK	320		

```
TOXCST_BioSeek CAS numbers    OK      320
TOXCST_Cellumen        CAS numbers    OK      320
TOXCST_CellzDirect     CAS numbers    OK      320
TOXCST_Gentronix       CAS numbers    OK      320
TOXCST_NCGC    CAS numbers    OK      320
TOXCST_Novascreen      CAS numbers    OK      320
TOXCST_Solidus CAS numbers    OK      320
TOXCST_ToxRefDB        CAS numbers    OK      320
ECBPRS (20090917)      CAS numbers,EINECS numbers    OK      80410
PubChem Structures + Assays (20090924)        CAS numbers    OK      1275
Leadscope_carc_level_2 CAS numbers    OK      600
Leadscope_ccris_genetox        CAS numbers    OK      1410
Leadscope_cder_chronic CAS numbers    OK      2
Leadscope_cder_genetox CAS numbers    OK      12
Leadscope_cder_genetox CAS numbers    ERROR   1
Leadscope_cder_repro_dev       CAS numbers    OK      2
Leadscope_cfsan_acute  CAS numbers    OK      690
Leadscope_cfsan_chronic        CAS numbers    OK      307
Leadscope_cfsan_genetox        CAS numbers    OK      247
Leadscope_cfsan_repro_dev      CAS numbers    OK      135
Leadscope_fda_marketed_drugs   CAS numbers    OK      451
Leadscope_genetox_level_2      CAS numbers    OK      1723
Leadscope_ntp_genetox  CAS numbers    OK      956
```

```
Datasets       Mode    label   Number of compounds
ECHA list of pre-registered substances (20090327),ChemIDplus (20110503),Chemical Identifier Resolver
(20110212),ChemDraw
(20110505),CPDBAS,DBPCAN,EPAFHM,FDAMDD,HPVCSI,HPVISD,IRISTR,KIERBL,NCTRER,NTPBSI,NTPHTS,ISSCAN,ISSMIC
,TOXCST,TOXCST_ACEA,TOXCST_Attagene,TOXCST_BioSeek,TOXCST_Cellumen,TOXCST_CellzDirect,TOXCST_Gentroni
x,TOXCST_NCGC,TOXCST_Novascreen,TOXCST_Solidus,TOXCST_ToxRefDB,TXCST2,ECETOC Technical Report No. 66
Skin irritation and corrosion Reference Chemicals data base (1995),Local Lymph Node Data for the
Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005),Local
Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1
2010),Bioconcentration factor (BCF) Gold Standard Database,Benchmark Data Set for In Silico
Prediction of Ames Mutagenicity,Bursi AMES Toxicity
Dataset,EPI_AOP,EPI_BCF,EPI_BioHC,EPI_Biowin,EPI_Boil_Pt,EPI_Henry,EPI_KM,EPI_KOA,EPI_Kowwin,EPI_Melt
_Pt,EPI_PCKOC,EPI_VP,EPI_WaterFrag,EPI_Wskowwin,ECBPRS      CAS numbers,EINECS numbers    OK
       168285
ECHA list of pre-registered substances (20090327),ECETOC Technical Report No. 66 Skin irritation and
corrosion Reference Chemicals data base (1995),Bioconcentration factor (BCF) Gold Standard
Database,EPI_BCF,EPI_Biowin,EPI_Boil_Pt,EPI_Melt_Pt,EPI_Wskowwin,Leadscope_cder_genetox   EINECS
numbers,CAS numbers    ERROR   196
```

```
user_name      label   text    count(*)
comparison     OK      Comparison between different sources  317727
comparison     ProbablyOK      Comparison between different sources  57146
comparison     Unknown Comparison between different sources  12735
comparison     Unknown single source  498950
comparison     ProbablyERROR  Comparison between different sources  21380
```

```
Dataset Structure type Number of compounds
ECHA list of pre-registered substances (20090327)    NA      143835
ChemIDplus (20110503)  MARKUSH 2964
ChemIDplus (20110503)  SMILES  124
ChemIDplus (20110503)  2D no H 72486
ChemIDplus (20110503)  2D with H       4894
Chemical Identifier Resolver (20110212)       SMILES  100
Chemical Identifier Resolver (20110212)       2D no H 2832
Chemical Identifier Resolver (20110212)       2D with H       70053
ChemDraw (20110505)    SMILES  2
ChemDraw (20110505)    2D no H 21335
ChemDraw (20110505)    2D with H       1182
CPDBAS  NA      39
CPDBAS  SMILES  4
CPDBAS  2D no H 1418
CPDBAS  2D with H       86
DBPCAN  2D no H 209
EPAFHM  2D no H 606
EPAFHM  2D with H       11
FDAMDD  2D no H 984
```

FDAMDD 2D with H 232
HPVCSI NA 1068
HPVCSI SMILES 9
HPVCSI 2D no H 2454
HPVCSI 2D with H 17
HPVISD NA 302
HPVISD 2D no H 696
HPVISD 2D with H 8
IRISTR NA 5
IRISTR SMILES 25
IRISTR 2D no H 507
IRISTR 2D with H 7
KIERBL 2D no H 276
KIERBL 2D with H 2
NCTRER 2D no H 200
NCTRER 2D with H 32
NTPBSI NA 151
NTPBSI SMILES 4
NTPBSI 2D no H 2109
NTPBSI 2D with H 66
NTPHTS NA 14
NTPHTS 2D no H 1342
NTPHTS 2D with H 52
ISSCAN NA 12
ISSCAN 2D no H 1090
ISSCAN 2D with H 11
ISSCAN 3D no H 35
ISSCAN 3D with H 2
ISSMIC NA 6
ISSMIC 2D no H 124
ISSMIC 2D with H 21
ISSSTY NA 1
ISSSTY 2D no H 9
ISSSTY 2D with H 213
TOXCST 2D no H 311
TOXCST 2D with H 9
TXCST2 2D no H 931
TXCST2 2D with H 29
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) NA 1
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) 2D no H 171
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) 2D with H 4
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) SMILES 2
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) 2D no H 1
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) 2D with H 16
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) 3D no H 1
Local Lymph Node Data for the Evaluation of Skin Sensitization - Compilation of historical data (Dermatitis Vol 16 No 4 2005) 3D with H 189
Local Lymph Node Data for the Evaluation of Skin Sensitization - Second compilation (Dermatitis Vol 21 No 1 2010) SMILES 108
Bioconcentration factor (BCF) Gold Standard Database NA 1130
Benchmark Data Set for pKa Prediction of Monoprotic Small Molecules the SMARTS Way SMILES 185
Benchmark Data Set for In Silico Prediction of Ames Mutagenicity 2D no H 6512
Bursi AMES Toxicity Dataset 2D no H 1
Bursi AMES Toxicity Dataset 2D with H 127
Bursi AMES Toxicity Dataset 3D no H 21
Bursi AMES Toxicity Dataset 3D with H 4188
EPI_AOP 2D no H 810
EPI_AOP 2D with H 8
EPI_BCF 2D no H 679
EPI_BCF 2D with H 6
EPI_BioHC 2D no H 175
EPI_Biowin 2D no H 1248
EPI_Biowin 2D with H 14
EPI_Biowin 3D no H 1
EPI_Boil_Pt NA 1
EPI_Boil_Pt 2D no H 5870
EPI_Boil_Pt 2D with H 18

EPI_Boil_Pt	3D no H	1	
EPI_Henry	2D no H	1813	
EPI_Henry	2D with H	16	
EPI_KM	2D no H	626	
EPI_KM	2D with H	5	
EPI_KOA	2D no H	301	
EPI_KOA	2D with H	7	
EPI_Kowwin	SMILES	7	
EPI_Kowwin	2D no H	15727	
EPI_Kowwin	2D with H	72	
EPI_Kowwin	3D no H	3	
EPI_Melt_Pt	NA	5	
EPI_Melt_Pt	SMILES	7	
EPI_Melt_Pt	2D no H	9881	
EPI_Melt_Pt	2D with H	157	
EPI_Melt_Pt	3D no H	1	
EPI_PCKOC	NA	2	
EPI_PCKOC	2D no H	781	
EPI_PCKOC	2D with H	5	
EPI_VP	NA	1	
EPI_VP	MARKUSH	1	
EPI_VP	SMILES	2	
EPI_VP	2D no H	3014	
EPI_VP	2D with H	19	
EPI_WaterFrag	NA	3	
EPI_WaterFrag	SMILES	2	
EPI_WaterFrag	2D no H	5687	
EPI_WaterFrag	2D with H	72	
EPI_Wskowwin	NA	3	
EPI_Wskowwin	2D no H	2311	
EPI_Wskowwin	2D with H	34	
TOXCST_ACEA	2D no H	311	
TOXCST_ACEA	2D with H	9	
TOXCST_Attagene	2D no H	311	
TOXCST_Attagene	2D with H	9	
TOXCST_BioSeek	2D no H	311	
TOXCST_BioSeek	2D with H	9	
TOXCST_Cellumen	2D no H	311	
TOXCST_Cellumen	2D with H	9	
TOXCST_CellzDirect	2D no H	311	
TOXCST_CellzDirect	2D with H	9	
TOXCST_Gentronix	2D no H	311	
TOXCST_Gentronix	2D with H	9	
TOXCST_NCGC	2D no H	311	
TOXCST_NCGC	2D with H	9	
TOXCST_Novascreen	2D no H	311	
TOXCST_Novascreen	2D with H	9	
TOXCST_Solidus	2D no H	311	
TOXCST_Solidus	2D with H	9	
TOXCST_ToxRefDB	2D no H	311	
TOXCST_ToxRefDB	2D with H	9	
ECBPRS (20090917)	MARKUSH	87	
ECBPRS (20090917)	SMILES	272	
ECBPRS (20090917)	2D no H	77237	
ECBPRS (20090917)	2D with H	2814	
NAME2STRUCTURE	NA	16	
NAME2STRUCTURE	SMILES	70630	
PubChem Structures + Assays (20090924)	NA	45	
PubChem Structures + Assays (20090924)	SMILES	164	
PubChem Structures + Assays (20090924)	2D no H	471094	
PubChem Structures + Assays (20090924)	2D with H	2643	
PubChem Structures + Assays (20090924)	3D no H	15	
PubChem Structures + Assays (20090924)	3D with H	4	
Leadscope_carc_level_2	NA	141	
Leadscope_carc_level_2	SMILES	2847	
Leadscope_ccris_genetox	NA	559	
Leadscope_ccris_genetox	SMILES	7442	
Leadscope_cder_chronic	NA	4	
Leadscope_cder_chronic	SMILES	117	
Leadscope_cder_genetox	NA	242	
Leadscope_cder_genetox	SMILES	94	
Leadscope_cder_repro_dev	SMILES	58	
Leadscope_cfsan_acute	NA	100	
Leadscope_cfsan_acute	SMILES	970	

Leadscope_cfsan_chronic	NA	106
Leadscope_cfsan_chronic	SMILES	549
Leadscope_cfsan_genetox	NA	130
Leadscope_cfsan_genetox	SMILES	566
Leadscope_cfsan_repro_dev	NA	47
Leadscope_cfsan_repro_dev	SMILES	265
Leadscope_fda_marketed_drugs	NA	321
Leadscope_fda_marketed_drugs	SMILES	6316
Leadscope_genetox_level_2	NA	748
Leadscope_genetox_level_2	SMILES	9407
Leadscope_ntp_genetox	NA	71
Leadscope_ntp_genetox	SMILES	2057
Pharmatrope_AERS_hepatobiliary_system	SMILES	1274

Appendix B: OpenTox final database structure

A full SQL dump of the final OpenTox database structure is provided below:

```
-- MySQL dump 10.13 Distrib 5.5.12, for Linux (x86_64)
--
-- Host: localhost Database: ambit2
-- ------------------------------------------------------
-- Server version 5.5.12-log
/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!40101 SET NAMES utf8 */;
/*!40103 SET @OLD_TIME_ZONE=@@TIME_ZONE */;
/*!40103 SET TIME_ZONE='+00:00' */;
/*!40014 SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0 */;
/*!40014 SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0 */;
/*!40101 SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='NO_AUTO_VALUE_ON_ZERO' */;
/*!40111 SET @OLD_SQL_NOTES=@@SQL_NOTES, SQL_NOTES=0 */;
--
-- Current Database: `ambit2`
--
/*!40000 DROP DATABASE IF EXISTS `ambit2`*/;
CREATE DATABASE /*!32312 IF NOT EXISTS*/ `ambit2` /*!40100 DEFAULT CHARACTER SET utf8
COLLATE utf8_bin */;
USE `ambit2`;
--
-- Table structure for table `atom_distance`
--
DROP TABLE IF EXISTS `atom_distance`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `atom_distance` (
`iddistance` int(10) unsigned NOT NULL AUTO_INCREMENT,
`atom1` varchar(2) COLLATE utf8_bin NOT NULL DEFAULT 'C',
`atom2` varchar(2) COLLATE utf8_bin NOT NULL DEFAULT 'C',
`distance` int(10) NOT NULL DEFAULT '0',
PRIMARY KEY (`iddistance`),
UNIQUE KEY `atom1` (`atom1`,`atom2`,`distance`),
KEY `distance` (`distance`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `atom_structure`
--
DROP TABLE IF EXISTS `atom_structure`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `atom_structure` (
`idstructure` int(11) unsigned NOT NULL DEFAULT '0',
`iddistance` int(11) unsigned NOT NULL AUTO_INCREMENT,
PRIMARY KEY (`iddistance`,`idstructure`),
KEY `adistance` (`idstructure`),
CONSTRAINT `atom_distance_fk_1` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `atom_distance_fk_2` FOREIGN KEY (`iddistance`) REFERENCES `atom_distance`
(`iddistance`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `bookmark`
--
DROP TABLE IF EXISTS `bookmark`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `bookmark` (
`idbookmark` int(10) unsigned NOT NULL AUTO_INCREMENT,
`creator` varchar(45) COLLATE utf8_bin NOT NULL COMMENT 'dc:creator',
`recalls` text COLLATE utf8_bin NOT NULL COMMENT 'b:recalls Relates the bookmark with
the resource that has been bookmarked. ',
`hasTopic` varchar(255) COLLATE utf8_bin NOT NULL COMMENT 'b:hasTopic Associates the
bookmark with a Topic ',
`title` varchar(45) COLLATE utf8_bin NOT NULL COMMENT 'dc:title',
`description` text COLLATE utf8_bin NOT NULL COMMENT 'dc:description',
```

```
`created` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP
COMMENT 'a:created The date and time on which the bookmark was created. Format should be
YYYY-MM-DDTHH:MM[:SS]TZD (see [DATETIME])',
`date` timestamp NOT NULL DEFAULT '2009-12-31 23:01:01' COMMENT 'dc:date The date and
time on which the bookmark was last modified. Format should be YYYY-MM-DDTHH:MM[:SS]TZD
(see [DATETIME])',
PRIMARY KEY (`idbookmark`),
KEY `Index_4` (`creator`,`hasTopic`,`date`) USING BTREE,
KEY `Index_3` (`hasTopic`) USING BTREE,
KEY `Index_2` (`creator`,`hasTopic`,`title`) USING BTREE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `catalog_references`
--
DROP TABLE IF EXISTS `catalog_references`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `catalog_references` (
`idreference` int(11) unsigned NOT NULL AUTO_INCREMENT,
`title` varchar(255) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
`url` varchar(255) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
`type`
enum('Unknown','Dataset','Algorithm','Model','BibtexEntry','BibtexArticle','BibtexBook',
'Feature') NOT NULL DEFAULT 'Dataset',
PRIMARY KEY (`idreference`),
UNIQUE KEY `Index_2` (`title`)
) ENGINE=InnoDB AUTO_INCREMENT=221919 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `chem_relation`
--
DROP TABLE IF EXISTS `chem_relation`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `chem_relation` (
`idchemical1` int(10) unsigned NOT NULL AUTO_INCREMENT,
`idchemical2` int(10) unsigned NOT NULL,
`relation` varchar(64) NOT NULL,
PRIMARY KEY (`idchemical1`,`idchemical2`,`relation`),
KEY `FK_chem_relation_2` (`idchemical2`),
CONSTRAINT `FK_chem_relation_1` FOREIGN KEY (`idchemical1`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_chem_relation_2` FOREIGN KEY (`idchemical2`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `chemicals`
--
DROP TABLE IF EXISTS `chemicals`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `chemicals` (
`idchemical` int(11) unsigned NOT NULL AUTO_INCREMENT,
`inchi` text CHARACTER SET latin1 COLLATE latin1_bin,
`smiles` text CHARACTER SET latin1 COLLATE latin1_bin,
`formula` varchar(64) DEFAULT NULL,
`inchikey` varchar(27) DEFAULT '0',
`label` enum('OK','UNKNOWN','ERROR') NOT NULL DEFAULT 'UNKNOWN',
PRIMARY KEY (`idchemical`),
KEY `index_smiles` (`smiles`(760)) USING BTREE,
KEY `index_idchemical` (`idchemical`) USING BTREE,
KEY `index_inchi` (`inchi`(767)) USING BTREE,
KEY `index_formula` (`formula`) USING BTREE,
KEY `index_inchikey` (`inchikey`) USING BTREE,
KEY `index_label` (`label`) USING BTREE
) ENGINE=InnoDB AUTO_INCREMENT=646043 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `dictionary`
--
DROP TABLE IF EXISTS `dictionary`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
```

```
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `dictionary` (
`idsubject` int(10) unsigned NOT NULL,
`relationship` enum('is_a','is_part_of') COLLATE utf8_bin NOT NULL DEFAULT 'is_a',
`idobject` int(10) unsigned NOT NULL,
PRIMARY KEY (`idsubject`,`relationship`,`idobject`),
KEY `FK_dictionary_2` (`idobject`),
CONSTRAINT `FK_dictionary_1` FOREIGN KEY (`idsubject`) REFERENCES `template`
(`idtemplate`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_dictionary_2` FOREIGN KEY (`idobject`) REFERENCES `template`
(`idtemplate`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `fp1024`
--
DROP TABLE IF EXISTS `fp1024`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `fp1024` (
`idchemical` int(10) unsigned NOT NULL DEFAULT '0',
`fp1` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp2` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp3` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp4` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp5` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp6` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp7` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp8` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp9` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp10` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp11` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp12` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp13` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp14` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp15` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp16` bigint(20) unsigned NOT NULL DEFAULT '0',
`time` int(10) unsigned DEFAULT '0',
`bc` int(6) NOT NULL DEFAULT '0',
`status` enum('invalid','valid','error') COLLATE utf8_bin NOT NULL DEFAULT 'invalid',
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`version` int(10) unsigned zerofill NOT NULL DEFAULT '0000000000',
PRIMARY KEY (`idchemical`),
KEY `fpall`
(`fp1`,`fp2`,`fp3`,`fp4`,`fp5`,`fp6`,`fp7`,`fp8`,`fp9`,`fp10`,`fp11`,`fp12`,`fp13`,`fp14
`,`fp15`,`fp16`),
KEY `time` (`time`),
KEY `status` (`status`),
CONSTRAINT `fp1024_ibfk_1` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `fp1024_struc`
--
DROP TABLE IF EXISTS `fp1024_struc`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `fp1024_struc` (
`idchemical` int(10) unsigned NOT NULL DEFAULT '0',
`idstructure` int(10) unsigned NOT NULL DEFAULT '0',
`fp1` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp2` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp3` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp4` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp5` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp6` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp7` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp8` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp9` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp10` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp11` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp12` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp13` bigint(20) unsigned NOT NULL DEFAULT '0',
```

```
`fp14` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp15` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp16` bigint(20) unsigned NOT NULL DEFAULT '0',
`time` int(10) unsigned DEFAULT '0',
`bc` int(6) NOT NULL DEFAULT '0',
`status` enum('invalid','valid','error') COLLATE utf8_bin NOT NULL DEFAULT 'invalid',
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`version` int(10) unsigned zerofill NOT NULL DEFAULT '0000000000',
PRIMARY KEY (`idchemical`,`idstructure`) USING BTREE,
KEY `fpall`
(`fp1`,`fp2`,`fp3`,`fp4`,`fp5`,`fp6`,`fp7`,`fp8`,`fp9`,`fp10`,`fp11`,`fp12`,`fp13`,`fp14
`,`fp15`,`fp16`),
KEY `time` (`time`),
KEY `status` (`status`),
KEY `fp1024struc_ibfk_2` (`idstructure`),
CONSTRAINT `fp1024struc_ibfk_1` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `fp1024struc_ibfk_2` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `fpae`
--
DROP TABLE IF EXISTS `fpae`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `fpae` (
`idfpae` int(10) unsigned NOT NULL AUTO_INCREMENT,
`ae` varchar(255) CHARACTER SET latin1 NOT NULL DEFAULT '',
PRIMARY KEY (`idfpae`),
UNIQUE KEY `ae` (`ae`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1 COLLATE=latin1_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `fpaechemicals`
--
DROP TABLE IF EXISTS `fpaechemicals`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `fpaechemicals` (
`idchemical` int(10) unsigned NOT NULL DEFAULT '0',
`idfpae1` int(10) unsigned NOT NULL,
`idfpae2` int(10) unsigned NOT NULL,
`idfpae3` int(10) unsigned NOT NULL,
`idfpae4` int(10) unsigned NOT NULL,
`idfpae5` int(10) unsigned NOT NULL,
`idfpae6` int(10) unsigned NOT NULL,
`status` enum('valid','invalid','error') CHARACTER SET utf8 COLLATE utf8_bin NOT NULL
DEFAULT 'valid',
`freq` int(10) unsigned NOT NULL DEFAULT '1',
`atom` varchar(6) COLLATE latin1_bin NOT NULL,
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP,
PRIMARY KEY (`idchemical`,`atom`) USING BTREE,
KEY `FK_fpaelevels_1` (`idfpae1`),
KEY `FK_fpaelevels_2` (`idfpae2`),
KEY `FK_fpaelevels_3` (`idfpae3`),
KEY `FK_fpaelevels_4` (`idfpae4`),
KEY `FK_fpaelevels_5` (`idfpae5`),
KEY `FK_fpaelevels_6` (`idfpae6`),
KEY `Index_8` (`status`),
KEY `Index_9` (`atom`,`idfpae1`,`idfpae2`,`idfpae3`,`idfpae4`,`idfpae5`,`idfpae6`),
CONSTRAINT `FK_fpaelevels_1` FOREIGN KEY (`idfpae1`) REFERENCES `fpae` (`idfpae`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_fpaelevels_2` FOREIGN KEY (`idfpae2`) REFERENCES `fpae` (`idfpae`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_fpaelevels_3` FOREIGN KEY (`idfpae3`) REFERENCES `fpae` (`idfpae`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_fpaelevels_4` FOREIGN KEY (`idfpae4`) REFERENCES `fpae` (`idfpae`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_fpaelevels_5` FOREIGN KEY (`idfpae5`) REFERENCES `fpae` (`idfpae`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_fpaelevels_6` FOREIGN KEY (`idfpae6`) REFERENCES `fpae` (`idfpae`) ON
DELETE CASCADE ON UPDATE CASCADE
```

```
) ENGINE=InnoDB DEFAULT CHARSET=latin1 COLLATE=latin1_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `funcgroups`
--
DROP TABLE IF EXISTS `funcgroups`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `funcgroups` (
`idfuncgroup` int(10) unsigned NOT NULL,
`name` varchar(45) COLLATE utf8_bin NOT NULL,
`smarts` blob NOT NULL,
`user_name` varchar(16) COLLATE utf8_bin DEFAULT NULL,
PRIMARY KEY (`idfuncgroup`),
UNIQUE KEY `Index_2` (`name`),
KEY `FK_funcgroups_1` (`user_name`),
CONSTRAINT `FK_funcgroups_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`)
ON DELETE SET NULL ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `models`
--
DROP TABLE IF EXISTS `models`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `models` (
`idmodel` int(10) unsigned NOT NULL AUTO_INCREMENT,
`name` varchar(255) COLLATE utf8_bin NOT NULL,
`idquery` int(10) unsigned DEFAULT NULL COMMENT 'dataset',
`predictors` int(10) unsigned NOT NULL COMMENT 'template for predictors',
`dependent` int(10) unsigned NOT NULL COMMENT 'template for dependent variables',
`content` longblob NOT NULL,
`algorithm` varchar(255) COLLATE utf8_bin NOT NULL DEFAULT 'N/A' COMMENT 'URI of the
algorithm',
`mediatype` varchar(48) COLLATE utf8_bin NOT NULL DEFAULT 'application/java' COMMENT
'Content formats: JAVA_CLASS, WEKA_BASE64, PMML',
`parameters` text COLLATE utf8_bin COMMENT 'Model parameters',
`predicted` int(10) unsigned NOT NULL COMMENT 'template for predicted variables',
`hidden` tinyint(1) NOT NULL DEFAULT '0',
`creator` varchar(45) COLLATE utf8_bin NOT NULL DEFAULT 'guest',
`dataset` varchar(255) COLLATE utf8_bin DEFAULT NULL COMMENT 'dataset uri',
`user_name` varchar(16) COLLATE utf8_bin NOT NULL DEFAULT 'guest',
PRIMARY KEY (`idmodel`),
UNIQUE KEY `Index_5` (`name`) USING BTREE,
KEY `FK_models_predictors` (`predictors`),
KEY `FK_models_dataset` (`idquery`),
KEY `FK_models_dependent` (`dependent`),
KEY `Index_6` (`algorithm`),
KEY `Index_7` (`parameters`(255)),
KEY `FK_models_predicted` (`predicted`),
KEY `Index_creator` (`creator`),
KEY `Index_10` (`dataset`),
KEY `FK_models_users` (`user_name`),
CONSTRAINT `FK_models_dataset` FOREIGN KEY (`idquery`) REFERENCES `query` (`idquery`) ON
UPDATE CASCADE,
CONSTRAINT `FK_models_dependent` FOREIGN KEY (`dependent`) REFERENCES `template`
(`idtemplate`) ON UPDATE CASCADE,
CONSTRAINT `FK_models_predicted` FOREIGN KEY (`predicted`) REFERENCES `template`
(`idtemplate`) ON UPDATE CASCADE,
CONSTRAINT `FK_models_predictors` FOREIGN KEY (`predictors`) REFERENCES `template`
(`idtemplate`) ON UPDATE CASCADE,
CONSTRAINT `FK_models_users` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`)
) ENGINE=InnoDB AUTO_INCREMENT=29580 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Temporary table structure for view `ontology`
--
DROP TABLE IF EXISTS `ontology`;
/*!50001 DROP VIEW IF EXISTS `ontology`*/;
SET @saved_cs_client = @@character_set_client;
SET character_set_client = utf8;
/*!50001 CREATE TABLE `ontology` (
`subjectid` int(10) unsigned,
```

```
`objectid` int(10) unsigned,
`subject` varchar(255),
`relationship` enum('is_a','is_part_of'),
`object` varchar(255)
) ENGINE=MyISAM */;
SET character_set_client = @saved_cs_client;
--
-- Table structure for table `properties`
--
DROP TABLE IF EXISTS `properties`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `properties` (
`idproperty` int(10) unsigned NOT NULL AUTO_INCREMENT,
`idreference` int(11) unsigned NOT NULL DEFAULT '0',
`name` varchar(255) COLLATE utf8_bin NOT NULL DEFAULT '',
`units` varchar(16) COLLATE utf8_bin NOT NULL DEFAULT '',
`comments` varchar(255) COLLATE utf8_bin NOT NULL DEFAULT '',
`islocal` tinyint(1) NOT NULL DEFAULT '0',
`ptype` set('STRING','NUMERIC') COLLATE utf8_bin DEFAULT NULL,
PRIMARY KEY (`idproperty`) USING BTREE,
UNIQUE KEY `ddictionary_name` (`name`,`idreference`) USING BTREE,
KEY `ddictionary_idref` (`idreference`),
CONSTRAINT `FK_properties_1` FOREIGN KEY (`idreference`) REFERENCES `catalog_references`
(`idreference`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=218593 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `property_annotation`
--
DROP TABLE IF EXISTS `property_annotation`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `property_annotation` (
`idproperty` int(10) unsigned NOT NULL AUTO_INCREMENT COMMENT 'subject',
`rdf_type` varchar(45) COLLATE utf8_unicode_ci NOT NULL DEFAULT 'Feature',
`predicate` varchar(45) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL COMMENT 'property',
`object` text CHARACTER SET utf8 NOT NULL COMMENT 'object',
PRIMARY KEY (`idproperty`,`rdf_type`,`predicate`,`object`(250)) USING BTREE,
KEY `Index_2` (`predicate`,`object`(250)) USING BTREE,
CONSTRAINT `FK_property_annotation_1` FOREIGN KEY (`idproperty`) REFERENCES `properties`
(`idproperty`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_unicode_ci;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `property_ci`
--
DROP TABLE IF EXISTS `property_ci`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `property_ci` (
`id_ci` int(10) unsigned NOT NULL AUTO_INCREMENT,
`value_ci` varchar(255) NOT NULL,
PRIMARY KEY (`id_ci`),
UNIQUE KEY `Index_3` (`value_ci`)
) ENGINE=InnoDB AUTO_INCREMENT=2758241 DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `property_pairstruc`
--
DROP TABLE IF EXISTS `property_pairstruc`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `property_pairstruc` (
`idstructure1` int(10) unsigned NOT NULL AUTO_INCREMENT COMMENT 'First structure id',
`idstructure2` int(10) unsigned NOT NULL COMMENT 'Second structure id',
`idproperty` int(10) unsigned NOT NULL COMMENT 'Property id',
`user_name` varchar(16) COLLATE utf8_bin NOT NULL COMMENT 'User',
`status` enum('OK','UNKNOWN','ERROR','TRUNCATED') COLLATE utf8_bin NOT NULL,
`text` text COLLATE utf8_bin COMMENT 'Text value, if longer than allowed by
property_string',
`idvalue_string` int(10) unsigned NOT NULL COMMENT 'link to property_string',
`value_num` double NOT NULL COMMENT 'numeric value',
`idtype` enum('STRING','NUMERIC') COLLATE utf8_bin NOT NULL,
```

```
PRIMARY KEY (`idstructure1`,`idstructure2`,`idproperty`) USING BTREE,
KEY `FK_relationship_struc_2` (`idstructure2`),
KEY `FK_relationship_struc_3` (`idproperty`),
KEY `FK_relationship_struc_4` (`user_name`),
KEY `FK_relationship_struc_5` (`idvalue_string`),
CONSTRAINT `FK_relationship_struc_1` FOREIGN KEY (`idstructure1`) REFERENCES `structure`
(`idstructure`),
CONSTRAINT `FK_relationship_struc_2` FOREIGN KEY (`idstructure2`) REFERENCES `structure`
(`idstructure`),
CONSTRAINT `FK_relationship_struc_3` FOREIGN KEY (`idproperty`) REFERENCES `properties`
(`idproperty`),
CONSTRAINT `FK_relationship_struc_4` FOREIGN KEY (`user_name`) REFERENCES `users`
(`user_name`),
CONSTRAINT `FK_relationship_struc_5` FOREIGN KEY (`idvalue_string`) REFERENCES
`property_string` (`idvalue_string`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `property_string`
--
DROP TABLE IF EXISTS `property_string`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `property_string` (
`idvalue_string` int(10) unsigned NOT NULL AUTO_INCREMENT,
`value` varchar(255) COLLATE utf8_bin NOT NULL,
PRIMARY KEY (`idvalue_string`),
UNIQUE KEY `Index_3` (`value`)
) ENGINE=InnoDB AUTO_INCREMENT=127836548 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50017 DEFINER=`root`@`127.0.0.1`*/ /*!50003 TRIGGER insert_string_ci
AFTER INSERT ON property_string
FOR EACH ROW BEGIN
INSERT IGNORE INTO property_ci (value_ci) values (NEW.value);
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50017 DEFINER=`root`@`127.0.0.1`*/ /*!50003 TRIGGER update_string_ci
AFTER UPDATE ON property_string
FOR EACH ROW BEGIN
UPDATE property_ci set value_ci=NEW.value where value_ci=OLD.value;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
--
-- Table structure for table `property_tuples`
```

```
--
DROP TABLE IF EXISTS `property_tuples`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `property_tuples` (
`idtuple` int(10) unsigned NOT NULL,
`id` int(10) unsigned NOT NULL,
PRIMARY KEY (`idtuple`,`id`) USING BTREE,
KEY `FK_property_tuples_2` (`id`),
CONSTRAINT `FK_property_tuples_2` FOREIGN KEY (`id`) REFERENCES `property_values` (`id`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_property_tuple_1` FOREIGN KEY (`idtuple`) REFERENCES `tuples` (`idtuple`)
ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50017 DEFINER=`root`@`127.0.0.1`*/ /*!50003 TRIGGER
insert_property_tuple AFTER INSERT ON property_tuples
FOR EACH ROW BEGIN
INSERT IGNORE INTO template_def (idtemplate,idproperty,`order`) (
SELECT idtemplate,idproperty,idproperty FROM
(SELECT idtemplate FROM src_dataset join tuples using(id_srcdataset) WHERE
idtuple=NEW.idtuple) a
JOIN
(SELECT idproperty from property_values WHERE id=NEW.id) b
) ;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
--
-- Table structure for table `property_values`
--
DROP TABLE IF EXISTS `property_values`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `property_values` (
`id` int(10) unsigned NOT NULL AUTO_INCREMENT,
`idproperty` int(10) unsigned NOT NULL,
`idstructure` int(10) unsigned NOT NULL,
`idchemical` int(10) unsigned NOT NULL,
`user_name` varchar(16) COLLATE utf8_bin NOT NULL,
`status` enum('OK','UNKNOWN','ERROR','TRUNCATED') COLLATE utf8_bin NOT NULL DEFAULT
'UNKNOWN',
`text` text COLLATE utf8_bin,
`idvalue_string` int(10) unsigned DEFAULT NULL,
`value_num` double(14,4) DEFAULT NULL,
`idtype` enum('STRING','NUMERIC') COLLATE utf8_bin NOT NULL DEFAULT 'STRING',
PRIMARY KEY (`id`),
UNIQUE KEY `Index_1` (`idproperty`,`idstructure`) USING BTREE,
KEY `FK_property_values_1` (`user_name`),
KEY `FK_property_values_2` (`idstructure`),
KEY `Index_2` (`value_num`),
KEY `FK_property_values_5` (`idvalue_string`),
KEY `Index_3` (`idproperty`,`idtype`) USING BTREE,
KEY `Index_8` (`idproperty`,`idvalue_string`),
KEY `Index_12` (`idchemical`,`idproperty`),
CONSTRAINT `FK_property_values_1` FOREIGN KEY (`user_name`) REFERENCES `users`
(`user_name`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_property_values_2` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_property_values_3` FOREIGN KEY (`idproperty`) REFERENCES `properties`
```

```
(`idproperty`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_property_values_5` FOREIGN KEY (`idvalue_string`) REFERENCES
`property_string` (`idvalue_string`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_property_values_6` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=132629422 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50017 DEFINER=`root`@`127.0.0.1`*/ /*!50003 TRIGGER
summary_chemical_prop_insert AFTER INSERT ON property_values
FOR EACH ROW BEGIN
UPDATE properties set ptype=CONCAT_WS(',',ptype,NEW.idtype) where
idproperty=NEW.idproperty;
INSERT IGNORE INTO summary_property_chemicals (idchemical,id_ci,idproperty)
SELECT idchemical,id_ci,idproperty from property_ci
JOIN structure
JOIN properties
WHERE
NEW.idvalue_string is not null
and value_ci = (select value from property_string where
idvalue_string=NEW.idvalue_string)
and idstructure=NEW.idstructure
and idproperty=NEW.idproperty;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50017 DEFINER=`root`@`127.0.0.1`*/ /*!50003 TRIGGER
summary_chemical_prop_update AFTER UPDATE ON property_values
FOR EACH ROW BEGIN
UPDATE properties set ptype=CONCAT_WS(',',ptype,NEW.idtype) where
idproperty=NEW.idproperty;
INSERT IGNORE INTO summary_property_chemicals (idchemical,id_ci,idproperty)
SELECT idchemical,id_ci,idproperty from property_ci
JOIN structure
JOIN properties
WHERE
NEW.idvalue_string is not null
and value_ci = (select value from property_string where
idvalue_string=NEW.idvalue_string)
and idstructure=NEW.idstructure
and idproperty=NEW.idproperty;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
--
-- Table structure for table `quality_chemicals`
--
```

```
DROP TABLE IF EXISTS `quality_chemicals`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `quality_chemicals` (
`idchemical` int(10) unsigned NOT NULL AUTO_INCREMENT,
`num_sources` int(10) unsigned NOT NULL,
`label` enum('Consensus','Majority','Unconfirmed','Ambiguous','Unknown') COLLATE
utf8_bin NOT NULL DEFAULT 'Unknown',
`num_structures` varchar(45) COLLATE utf8_bin NOT NULL DEFAULT '0',
`text` varchar(255) COLLATE utf8_bin NOT NULL,
PRIMARY KEY (`idchemical`),
KEY `Index_4` (`num_structures`),
KEY `Index_2` (`label`,`text`) USING BTREE
) ENGINE=InnoDB AUTO_INCREMENT=646043 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `quality_labels`
--
DROP TABLE IF EXISTS `quality_labels`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `quality_labels` (
`id` int(10) unsigned NOT NULL,
`user_name` varchar(16) COLLATE utf8_bin NOT NULL,
`label` enum('OK','ProbablyOK','Unknown','ProbablyERROR','ERROR') COLLATE utf8_bin NOT
NULL DEFAULT 'Unknown',
`text` text COLLATE utf8_bin,
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
PRIMARY KEY (`id`,`user_name`) USING BTREE,
KEY `FK_quality_labels_2` (`user_name`),
KEY `FK_quality_labels_3` (`label`),
CONSTRAINT `FK_quality_labels_1` FOREIGN KEY (`id`) REFERENCES `property_values` (`id`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_quality_labels_2` FOREIGN KEY (`user_name`) REFERENCES `users`
(`user_name`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `quality_pair`
--
DROP TABLE IF EXISTS `quality_pair`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `quality_pair` (
`idchemical` int(10) unsigned NOT NULL AUTO_INCREMENT,
`idstructure` int(10) unsigned NOT NULL,
`rel` int(10) unsigned NOT NULL DEFAULT '0' COMMENT 'number of same structures',
`user_name` varchar(16) COLLATE utf8_bin NOT NULL,
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`TEXT` text COLLATE utf8_bin,
PRIMARY KEY (`idchemical`,`idstructure`),
KEY `FK_qpair_1` (`user_name`),
KEY `FK_qpair_3` (`idstructure`),
KEY `Index_4` (`TEXT`(255)),
KEY `Index_5` (`idchemical`,`rel`) USING BTREE,
CONSTRAINT `FK_qpair_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_qpair_2` FOREIGN KEY (`idchemical`) REFERENCES `chemicals` (`idchemical`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_qpair_3` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=50342 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `quality_structure`
--
DROP TABLE IF EXISTS `quality_structure`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `quality_structure` (
`idstructure` int(10) unsigned NOT NULL,
`user_name` varchar(16) COLLATE utf8_bin NOT NULL,
`label` enum('OK','ProbablyOK','Unknown','ProbablyERROR','ERROR') COLLATE utf8_bin NOT
NULL DEFAULT 'Unknown',
```

```
`text` text COLLATE utf8_bin,
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
PRIMARY KEY (`idstructure`,`user_name`) USING BTREE,
KEY `FK_quality_struc_2` (`user_name`),
KEY `FK_quality_struc_3` (`label`),
CONSTRAINT `FK_quality_struc_1` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_quality_struc_2` FOREIGN KEY (`user_name`) REFERENCES `users`
(`user_name`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `query`
--
DROP TABLE IF EXISTS `query`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `query` (
`idquery` int(10) unsigned NOT NULL AUTO_INCREMENT,
`idsessions` int(10) unsigned NOT NULL,
`name` text COLLATE utf8_bin NOT NULL,
`content` text COLLATE utf8_bin NOT NULL,
`idtemplate` int(10) unsigned DEFAULT NULL,
PRIMARY KEY (`idquery`),
UNIQUE KEY `Index_3` (`name`(255),`idsessions`) USING BTREE,
KEY `FK_query_1` (`idsessions`),
KEY `FK_query_2` (`idtemplate`),
CONSTRAINT `FK_query_1` FOREIGN KEY (`idsessions`) REFERENCES `sessions` (`idsessions`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_query_2` FOREIGN KEY (`idtemplate`) REFERENCES `template` (`idtemplate`)
ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=17725 DEFAULT CHARSET=utf8 COLLATE=utf8_bin
COMMENT='Structure queries';
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `query_results`
--
DROP TABLE IF EXISTS `query_results`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `query_results` (
`idquery` int(10) unsigned NOT NULL,
`idchemical` int(10) unsigned NOT NULL,
`idstructure` int(11) unsigned NOT NULL,
`selected` tinyint(1) NOT NULL DEFAULT '1',
`metric` float(10,6) DEFAULT NULL,
`text` varchar(200) COLLATE utf8_bin DEFAULT NULL,
PRIMARY KEY (`idquery`,`idchemical`,`idstructure`),
KEY `FK_query_results_2` (`idstructure`),
KEY `FK_query_results_3` (`idchemical`),
KEY `Index_4` (`idquery`,`metric`) USING BTREE,
KEY `Index_5` (`idquery`),
KEY `Index_6` (`idquery`,`text`),
CONSTRAINT `FK_query_results_1` FOREIGN KEY (`idquery`) REFERENCES `query` (`idquery`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_query_results_2` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_query_results_3` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `roles`
--
DROP TABLE IF EXISTS `roles`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `roles` (
`role_name` varchar(16) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
PRIMARY KEY (`role_name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `sessions`
```

```
--
DROP TABLE IF EXISTS `sessions`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `sessions` (
`idsessions` int(10) unsigned NOT NULL AUTO_INCREMENT,
`user_name` varchar(16) COLLATE utf8_bin NOT NULL,
`started` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`completed` timestamp NOT NULL DEFAULT '2009-12-31 23:01:01',
`title` varchar(45) COLLATE utf8_bin NOT NULL DEFAULT 'temp',
PRIMARY KEY (`idsessions`),
UNIQUE KEY `Index_3` (`title`,`user_name`) USING BTREE,
KEY `FK_sessions_1` (`user_name`),
CONSTRAINT `FK_sessions_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`) ON
DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=18036 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `sk1024`
--
DROP TABLE IF EXISTS `sk1024`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `sk1024` (
`idchemical` int(10) unsigned NOT NULL DEFAULT '0',
`fp1` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp2` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp3` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp4` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp5` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp6` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp7` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp8` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp9` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp10` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp11` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp12` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp13` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp14` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp15` bigint(20) unsigned NOT NULL DEFAULT '0',
`fp16` bigint(20) unsigned NOT NULL DEFAULT '0',
`time` int(10) unsigned DEFAULT '0',
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`bc` int(6) NOT NULL DEFAULT '0',
`status` enum('invalid','valid','error') COLLATE utf8_bin NOT NULL DEFAULT 'invalid',
PRIMARY KEY (`idchemical`),
KEY `fpall`
(`fp1`,`fp2`,`fp3`,`fp4`,`fp5`,`fp6`,`fp7`,`fp8`,`fp9`,`fp10`,`fp11`,`fp12`,`fp13`,`fp14
`,`fp15`,`fp16`),
KEY `time` (`time`),
KEY `status` (`status`),
CONSTRAINT `sk1024_ibfk_1` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `src_dataset`
--
DROP TABLE IF EXISTS `src_dataset`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `src_dataset` (
`id_srcdataset` int(11) unsigned NOT NULL AUTO_INCREMENT,
`name` varchar(255) COLLATE utf8_bin NOT NULL DEFAULT 'default',
`user_name` varchar(16) COLLATE utf8_bin DEFAULT NULL,
`idreference` int(11) unsigned NOT NULL,
`created` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`idtemplate` int(10) unsigned DEFAULT NULL,
`licenseURI` varchar(128) COLLATE utf8_bin NOT NULL DEFAULT 'Unknown',
`rightsHolder` varchar(128) COLLATE utf8_bin NOT NULL DEFAULT 'Unknown',
`maintainer` varchar(45) COLLATE utf8_bin NOT NULL DEFAULT 'Unknown',
PRIMARY KEY (`id_srcdataset`),
UNIQUE KEY `src_dataset_name` (`name`),
KEY `FK_src_dataset_1` (`user_name`),
```

```
KEY `FK_src_dataset_2` (`idreference`),
KEY `FK_src_dataset_3` (`idtemplate`),
KEY `Index_6` (`maintainer`),
CONSTRAINT `FK_src_dataset_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`)
ON DELETE SET NULL ON UPDATE CASCADE,
CONSTRAINT `FK_src_dataset_2` FOREIGN KEY (`idreference`) REFERENCES
`catalog_references` (`idreference`) ON UPDATE CASCADE,
CONSTRAINT `FK_src_dataset_3` FOREIGN KEY (`idtemplate`) REFERENCES `template`
(`idtemplate`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=6515 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50017 DEFINER=`root`@`127.0.0.1`*/ /*!50003 TRIGGER
insert_dataset_template BEFORE INSERT ON src_dataset
FOR EACH ROW BEGIN
INSERT IGNORE INTO template (name) values (NEW.name);
SET NEW.idtemplate = (SELECT idtemplate FROM template where template.name=NEW.name);
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
--
-- Table structure for table `struc_dataset`
--
DROP TABLE IF EXISTS `struc_dataset`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `struc_dataset` (
`idstructure` int(10) unsigned NOT NULL,
`id_srcdataset` int(10) unsigned NOT NULL,
`created` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
PRIMARY KEY (`idstructure`,`id_srcdataset`),
KEY `struc_dataset` (`id_srcdataset`),
CONSTRAINT `struc_dataset_ibfk_1` FOREIGN KEY (`idstructure`) REFERENCES `structure`
(`idstructure`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `struc_dataset_ibfk_2` FOREIGN KEY (`id_srcdataset`) REFERENCES `src_dataset`
(`id_srcdataset`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `struc_fgroups`
--
DROP TABLE IF EXISTS `struc_fgroups`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `struc_fgroups` (
`idfuncgroup` int(10) unsigned NOT NULL,
`idchemical` int(10) unsigned NOT NULL DEFAULT '0',
PRIMARY KEY (`idfuncgroup`,`idchemical`) USING BTREE,
KEY `FK_struc_fgroups_1` (`idchemical`),
CONSTRAINT `FK_struc_fgroups_1` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_struc_fgroups_2` FOREIGN KEY (`idfuncgroup`) REFERENCES `funcgroups`
(`idfuncgroup`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `structure`
--
DROP TABLE IF EXISTS `structure`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
```

```
CREATE TABLE `structure` (
`idstructure` int(11) unsigned NOT NULL AUTO_INCREMENT,
`idchemical` int(11) unsigned NOT NULL,
`structure` blob NOT NULL,
`format` enum('SDF','CML','MOL','INC') COLLATE utf8_bin NOT NULL DEFAULT 'SDF',
`updated` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
`user_name` varchar(16) COLLATE utf8_bin DEFAULT NULL,
`type_structure` enum('NA','MARKUSH','SMILES','2D no H','2D with H','3D no H','3D with
H','optimized','experimental') COLLATE utf8_bin NOT NULL DEFAULT 'NA',
`label` enum('OK','UNKNOWN','ERROR') COLLATE utf8_bin NOT NULL DEFAULT 'UNKNOWN' COMMENT
'quality label',
`atomproperties` blob,
`preference` int(10) unsigned NOT NULL DEFAULT '9999',
PRIMARY KEY (`idstructure`),
KEY `FK_structure_2` (`user_name`),
KEY `idchemical` (`idchemical`) USING BTREE,
KEY `Index_4` (`label`),
KEY `Index_5` (`idstructure`,`user_name`),
KEY `Index_6` (`idchemical`,`preference`,`idstructure`) USING BTREE,
KEY `Index_pref` (`preference`,`idchemical`) USING BTREE,
CONSTRAINT `fk_idchemical` FOREIGN KEY (`idchemical`) REFERENCES `chemicals`
(`idchemical`) ON UPDATE CASCADE,
CONSTRAINT `FK_structure_2` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`)
ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=1064880 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `summary_property_chemicals`
--
DROP TABLE IF EXISTS `summary_property_chemicals`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `summary_property_chemicals` (
`idchemical` int(10) unsigned NOT NULL AUTO_INCREMENT,
`id_ci` int(10) unsigned NOT NULL,
`idproperty` int(10) unsigned NOT NULL,
PRIMARY KEY (`idchemical`,`id_ci`,`idproperty`),
KEY `FK_ppci_2` (`id_ci`),
KEY `FK_ppci_3` (`idproperty`),
KEY `Index_4` (`idchemical`,`idproperty`),
CONSTRAINT `FK_ppci_1` FOREIGN KEY (`idchemical`) REFERENCES `chemicals` (`idchemical`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_ppci_2` FOREIGN KEY (`id_ci`) REFERENCES `property_ci` (`id_ci`) ON
DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_ppci_3` FOREIGN KEY (`idproperty`) REFERENCES `properties` (`idproperty`)
ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=646043 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `template`
--
DROP TABLE IF EXISTS `template`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `template` (
`idtemplate` int(10) unsigned NOT NULL AUTO_INCREMENT,
`name` varchar(255) COLLATE utf8_bin DEFAULT NULL,
PRIMARY KEY (`idtemplate`),
UNIQUE KEY `template_list_index4157` (`name`) USING BTREE
) ENGINE=InnoDB AUTO_INCREMENT=184196 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `template_def`
--
DROP TABLE IF EXISTS `template_def`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `template_def` (
`idtemplate` int(10) unsigned NOT NULL,
`idproperty` int(10) unsigned NOT NULL,
`order` int(10) unsigned NOT NULL DEFAULT '0',
PRIMARY KEY (`idtemplate`,`idproperty`) USING BTREE,
KEY `FK_template_def_2` (`idproperty`),
CONSTRAINT `FK_template_def_1` FOREIGN KEY (`idtemplate`) REFERENCES `template`
```

```
(`idtemplate`) ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_template_def_2` FOREIGN KEY (`idproperty`) REFERENCES `properties`
(`idproperty`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Temporary table structure for view `template_properties`
--
DROP TABLE IF EXISTS `template_properties`;
/*!50001 DROP VIEW IF EXISTS `template_properties`*/;
SET @saved_cs_client = @@character_set_client;
SET character_set_client = utf8;
/*!50001 CREATE TABLE `template_properties` (
`idtemplate` int(10) unsigned,
`template` varchar(255),
`idproperty` int(10) unsigned,
`property` varchar(255),
`property_type` set('STRING','NUMERIC')
) ENGINE=MyISAM */;
SET character_set_client = @saved_cs_client;
--
-- Table structure for table `tuples`
--
DROP TABLE IF EXISTS `tuples`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `tuples` (
`idtuple` int(10) unsigned NOT NULL AUTO_INCREMENT,
`id_srcdataset` int(11) unsigned NOT NULL,
PRIMARY KEY (`idtuple`),
KEY `FK_tuples_1` (`id_srcdataset`),
CONSTRAINT `FK_tuples_1` FOREIGN KEY (`id_srcdataset`) REFERENCES `src_dataset`
(`id_srcdataset`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=1386773 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `user_roles`
--
DROP TABLE IF EXISTS `user_roles`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `user_roles` (
`user_name` varchar(16) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
`role_name` varchar(16) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
PRIMARY KEY (`user_name`,`role_name`),
KEY `FK_user_roles_2` (`role_name`),
CONSTRAINT `FK_user_roles_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`)
ON DELETE CASCADE ON UPDATE CASCADE,
CONSTRAINT `FK_user_roles_2` FOREIGN KEY (`role_name`) REFERENCES `roles` (`role_name`)
ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Table structure for table `users`
--
DROP TABLE IF EXISTS `users`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `users` (
`user_name` varchar(16) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
`password` varchar(45) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL,
`email` varchar(45) NOT NULL,
`registration_status` enum('commenced','confirmed','deleted') NOT NULL DEFAULT
'commenced',
`registration_date` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE
CURRENT_TIMESTAMP,
`registration_id` blob,
`title` varchar(6) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
`firstname` varchar(128) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
`lastname` varchar(128) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
`address` varchar(128) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
`country` varchar(128) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
`webpage` varchar(255) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
`affiliation` varchar(128) CHARACTER SET utf8 COLLATE utf8_bin NOT NULL DEFAULT '""',
```

```
`keywords` varchar(128) DEFAULT '"',
`reviewer` tinyint(1) NOT NULL DEFAULT '0' COMMENT 'true if wants to become a reviewer',
PRIMARY KEY (`user_name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Temporary table structure for view `values_all`
--
DROP TABLE IF EXISTS `values_all`;
/*!50001 DROP VIEW IF EXISTS `values_all`*/;
SET @saved_cs_client = @@character_set_client;
SET character_set_client = utf8;
/*!50001 CREATE TABLE `values_all` (
`idstructure` int(11) unsigned,
`idproperty` int(11) unsigned,
`name` varchar(255),
`value_string` varchar(255),
`value_number` double(14,4),
`idreference` int(11) unsigned
) ENGINE=MyISAM */;
SET character_set_client = @saved_cs_client;
--
-- Temporary table structure for view `values_number`
--
DROP TABLE IF EXISTS `values_number`;
/*!50001 DROP VIEW IF EXISTS `values_number`*/;
SET @saved_cs_client = @@character_set_client;
SET character_set_client = utf8;
/*!50001 CREATE TABLE `values_number` (
`id` int(10) unsigned,
`idproperty` int(10) unsigned,
`idstructure` int(10) unsigned,
`value` double(14,4),
`status` enum('OK','UNKNOWN','ERROR','TRUNCATED'),
`user_name` varchar(16)
) ENGINE=MyISAM */;
SET character_set_client = @saved_cs_client;
--
-- Temporary table structure for view `values_string`
--
DROP TABLE IF EXISTS `values_string`;
/*!50001 DROP VIEW IF EXISTS `values_string`*/;
SET @saved_cs_client = @@character_set_client;
SET character_set_client = utf8;
/*!50001 CREATE TABLE `values_string` (
`id` int(10) unsigned,
`idproperty` int(10) unsigned,
`idstructure` int(10) unsigned,
`value` text,
`status` enum('OK','UNKNOWN','ERROR','TRUNCATED'),
`user_name` varchar(16),
`name` varchar(255)
) ENGINE=MyISAM */;
SET character_set_client = @saved_cs_client;
--
-- Table structure for table `version`
--
DROP TABLE IF EXISTS `version`;
/*!40101 SET @saved_cs_client = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `version` (
`idmajor` int(5) unsigned NOT NULL,
`idminor` int(5) unsigned NOT NULL,
`date` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP,
`comment` varchar(45) COLLATE utf8_bin DEFAULT NULL,
PRIMARY KEY (`idmajor`,`idminor`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
/*!40101 SET character_set_client = @saved_cs_client */;
--
-- Dumping routines for database 'ambit2'
--
/*!50003 DROP FUNCTION IF EXISTS `sortstring` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
```

```
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 FUNCTION
`sortstring`(inString TEXT) RETURNS text CHARSET utf8
DETERMINISTIC
BEGIN
DECLARE delim CHAR(1) DEFAULT ',';
DECLARE strings INT DEFAULT 0;
DECLARE forward INT DEFAULT 1;
DECLARE backward INT;
DECLARE remain TEXT;
DECLARE swap1 TEXT;
DECLARE swap2 TEXT;
SET remain = inString;
SET backward = LOCATE(delim, remain);
WHILE backward != 0 DO
SET strings = strings + 1;
SET backward = LOCATE(delim, remain);
SET remain = SUBSTRING(remain, backward+1);
END WHILE;
IF strings < 2 THEN RETURN inString; END IF;
REPEAT
SET backward = strings;
REPEAT
SET swap1 = SUBSTRING_INDEX(SUBSTRING_INDEX(inString,delim,backward-1),delim,-1);
SET swap2 = SUBSTRING_INDEX(SUBSTRING_INDEX(inString,delim,backward),delim,-1);
IF swap1 > swap2 THEN
SET inString = TRIM(BOTH delim FROM CONCAT_WS(delim
,SUBSTRING_INDEX(inString,delim,backward-2)
,swap2,swap1
,SUBSTRING_INDEX(inString,delim,(backward-strings))));
END IF;
SET backward = backward - 1;
UNTIL backward < 2 END REPEAT;
SET forward = forward +1;
UNTIL forward + 1 > strings
END REPEAT;
RETURN inString;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 DROP FUNCTION IF EXISTS `sql_dataset_xtab` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 FUNCTION
`sql_dataset_xtab`(property_num VARCHAR(128),property_nom VARCHAR(128), dataset INT,bins
DOUBLE) RETURNS text CHARSET utf8
READS SQL DATA
begin
DECLARE x TEXT;
set @@group_concat_max_len=100000;
select concat(
'select a-mod(a,',bins,') \"',property_num,'\"\n'
, group_concat(distinct
concat(
', sum(','\n'
```

```
, ' if(b=\"',ifnull(text,value),'\"\n'
, ' , 1','\n'
, ' , 0','\n'
, ' )\n'
, ' )'
, ' "',value,'"\n'
)
order by ifnull(text,value)
separator ''
)
,',sum(if(b is null, 1, 0)) "N/A" '
,'from (','\n'
,' select a,b from','\n'
,' (','\n'
,' select value_num as a,idchemical from structure join struc_dataset
using(idstructure)','\n'
,' join property_values using(idstructure) join properties using(idproperty)','\n'
,' where name = "',property_num,'" and id_srcdataset=',dataset,' and value_num is not
null\n'
,' group by idchemical,value_num','\n'
,' ) as X','\n'
,' left join','\n'
,' (','\n'
,' select ifnull(text,value) as b,idchemical from structure join struc_dataset
using(idstructure)','\n'
,' join property_values using(idstructure) join property_string using(idvalue_string)
join properties using(idproperty)','\n'
,' where name = "',property_nom,'" and id_srcdataset=',dataset,'\n'
,' group by idchemical,value_num','\n'
,' ) as Y','\n'
,' using(idchemical)','\n'
,') as p','\n'
,'group by a-mod(a,',bins,')','\n'
) s
into @x
from struc_dataset join property_values using(idstructure) left join property_string
using(idvalue_string) join properties using(idproperty)
where name = property_nom and id_srcdataset=dataset;
return @x;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 DROP FUNCTION IF EXISTS `sql_xtab` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 FUNCTION
`sql_xtab`(property_num VARCHAR(128),property_nom VARCHAR(128), query INT,bins DOUBLE)
RETURNS text CHARSET utf8
READS SQL DATA
begin
set @x="";
set @@group_concat_max_len=100000;
select concat(
'select a-mod(a,',bins,') \"',property_num,'\"\n'
, group_concat(distinct
concat(
', sum(','\n'
, ' if(b=\"',value,'\"\n'
, ' , 1','\n'
, ' , 0','\n'
, ' )\n'
, ' )'
```

```
, ' "',value,'"\n'
)
order by value
separator ''
)
,',sum(if(b is null, 1, 0)) "N/A" '
,'from (','\n'
,' select a,b from','\n'
,' (','\n'
,' select value_num as a,idchemical from query_results','\n'
,' join property_values using(idstructure) join properties using(idproperty)','\n'
,' where name = "',property_num,'" and idquery=',query,' and value_num is not null\n'
,' group by idchemical,value_num','\n'
,' ) as X','\n'
,' left join','\n'
,' (','\n'
,' select value as b,idchemical from query_results','\n'
,' join property_values using(idstructure) join property_string using(idvalue_string)
join properties using(idproperty)','\n'
,' where name = "',property_nom,'" and idquery=',query,'\n'
,' group by idchemical,value_num','\n'
,' ) as Y','\n'
,' using(idchemical)','\n'
,') as p','\n'
,'group by a-mod(a,',bins,')','\n'
) s
into @x
from query_results join property_values using(idstructure) left join property_string
using(idvalue_string) join properties using(idproperty)
where name = property_nom and idquery=query;
return @x;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 DROP PROCEDURE IF EXISTS `copy_dataset_features` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 PROCEDURE
`copy_dataset_features`()
READS SQL DATA
BEGIN
DECLARE no_more_rows BOOLEAN;
DECLARE dataset_id INTEGER;
DECLARE dataset_name VARCHAR(255);
DECLARE template_id INTEGER;
DECLARE datasets CURSOR FOR
SELECT id_srcdataset,name,idtemplate FROM src_dataset ;
DECLARE CONTINUE HANDLER FOR NOT FOUND SET no_more_rows = TRUE;
SELECT "open";
OPEN datasets;
SELECT "start loop";
the_loop: LOOP
FETCH datasets into dataset_id,dataset_name,template_id;
IF no_more_rows THEN
CLOSE datasets;
LEAVE the_loop;
END IF;
SELECT dataset_name,dataset_id,template_id;
IF template_id IS NULL THEN
INSERT IGNORE INTO template (idtemplate,name) values (null,dataset_name);
UPDATE src_dataset, template SET src_dataset.idtemplate=template.idtemplate
WHERE id_srcdataset=dataset_id AND template.name=src_dataset.name;
```

```
ELSE
DELETE FROM template_def WHERE idtemplate= template_id;
END IF;
INSERT IGNORE into template_def (idtemplate,idproperty,`order`)
SELECT idtemplate,idproperty,idproperty
FROM property_values
JOIN property_tuples using(id)
JOIN tuples using (idtuple)
JOIN src_dataset using(id_srcdataset)
WHERE id_srcdataset=dataset_id
GROUP by idproperty;
END LOOP the_loop;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 DROP PROCEDURE IF EXISTS `p_dataset_xtab` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 PROCEDURE
`p_dataset_xtab`(IN property_num VARCHAR(128),property_nom VARCHAR(128),q INT,bins DOUBLE)
begin
set @x="";
select sql_dataset_xtab(property_num,property_nom,q,bins) into @x;
prepare xtab from @x;
execute xtab;
deallocate prepare xtab;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 DROP PROCEDURE IF EXISTS `p_xtab` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 PROCEDURE `p_xtab`(IN
property_num VARCHAR(128),property_nom VARCHAR(128),q INT,bins DOUBLE)
begin
set @x="";
select sql_xtab(property_num,property_nom,q,bins) into @x;
prepare xtab from @x;
execute xtab;
deallocate prepare xtab;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
/*!50003 DROP PROCEDURE IF EXISTS `setAtomEnvironment` */;
/*!50003 SET @saved_cs_client = @@character_set_client */ ;
/*!50003 SET @saved_cs_results = @@character_set_results */ ;
/*!50003 SET @saved_col_connection = @@collation_connection */ ;
```

```
/*!50003 SET character_set_client = utf8 */ ;
/*!50003 SET character_set_results = utf8 */ ;
/*!50003 SET collation_connection = utf8_general_ci */ ;
/*!50003 SET @saved_sql_mode = @@sql_mode */ ;
/*!50003 SET sql_mode =
'STRICT_TRANS_TABLES,STRICT_ALL_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FOR_DIVISION_BY_
ZERO,TRADITIONAL,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION' */ ;
DELIMITER ;;
/*!50003 CREATE*/ /*!50020 DEFINER=`root`@`127.0.0.1`*/ /*!50003 PROCEDURE
`setAtomEnvironment`(
IN chemical_id INTEGER,
IN chemical_atom VARCHAR(6),
IN ae_freq INT,
IN ae_1 VARCHAR(255),
IN ae_2 VARCHAR(255),
IN ae_3 VARCHAR(255),
IN ae_4 VARCHAR(255),
IN ae_5 VARCHAR(255),
IN ae_6 VARCHAR(255),
IN ae_status VARCHAR(10)
)
BEGIN
insert ignore into fpae (ae) values (ae_1),(ae_2),(ae_3),(ae_4),(ae_5),(ae_6);
insert into fpaechemicals
(idchemical,
atom,
freq,
idfpae1,
idfpae2,
idfpae3,
idfpae4,
idfpae5,
idfpae6,
`status`
)
SELECT chemical_id,
chemical_atom,
ae_freq,
a1.idfpae,
a2.idfpae,
a3.idfpae,
a4.idfpae,
a5.idfpae,
a6.idfpae,
ae_status
FROM fpae a1
join (SELECT idfpae FROM fpae where ae= ae_2) as a2
join (SELECT idfpae FROM fpae where ae= ae_3) as a3
join (SELECT idfpae FROM fpae where ae= ae_4) as a4
join (SELECT idfpae FROM fpae where ae= ae_5) as a5
join (SELECT idfpae FROM fpae where ae= ae_6) as a6
where a1.ae=ae_1
ON DUPLICATE KEY UPDATE
freq =values(freq),
idfpae1=values(idfpae1),
idfpae2=values(idfpae2),
idfpae3=values(idfpae3),
idfpae4=values(idfpae4),
idfpae5=values(idfpae5),
idfpae6=values(idfpae6),
`status`=values(`status`)
;
END */;;
DELIMITER ;
/*!50003 SET sql_mode = @saved_sql_mode */ ;
/*!50003 SET character_set_client = @saved_cs_client */ ;
/*!50003 SET character_set_results = @saved_cs_results */ ;
/*!50003 SET collation_connection = @saved_col_connection */ ;
--
-- Current Database: `ambit2`
--
USE `ambit2`;
--
-- Final view structure for view `ontology`
```

```
--
/*!50001 DROP TABLE IF EXISTS `ontology`*/;
/*!50001 DROP VIEW IF EXISTS `ontology`*/;
/*!50001 SET @saved_cs_client = @@character_set_client */;
/*!50001 SET @saved_cs_results = @@character_set_results */;
/*!50001 SET @saved_col_connection = @@collation_connection */;
/*!50001 SET character_set_client = utf8 */;
/*!50001 SET character_set_results = utf8 */;
/*!50001 SET collation_connection = utf8_general_ci */;
/*!50001 CREATE ALGORITHM=UNDEFINED */
/*!50013 DEFINER=`root`@`127.0.0.1` SQL SECURITY DEFINER */
/*!50001 VIEW `ontology` AS select `t1`.`idtemplate` AS `subjectid`,`t2`.`idtemplate` AS
`objectid`,`t1`.`name` AS `subject`,`d`.`relationship` AS `relationship`,`t2`.`name` AS
`object` from ((`template` `t1` join `dictionary` `d` on((`t1`.`idtemplate` =
`d`.`idsubject`))) join `template` `t2` on((`d`.`idobject` = `t2`.`idtemplate`))) */;
/*!50001 SET character_set_client = @saved_cs_client */;
/*!50001 SET character_set_results = @saved_cs_results */;
/*!50001 SET collation_connection = @saved_col_connection */;
--
-- Final view structure for view `template_properties`
--
/*!50001 DROP TABLE IF EXISTS `template_properties`*/;
/*!50001 DROP VIEW IF EXISTS `template_properties`*/;
/*!50001 SET @saved_cs_client = @@character_set_client */;
/*!50001 SET @saved_cs_results = @@character_set_results */;
/*!50001 SET @saved_col_connection = @@collation_connection */;
/*!50001 SET character_set_client = utf8 */;
/*!50001 SET character_set_results = utf8 */;
/*!50001 SET collation_connection = utf8_general_ci */;
/*!50001 CREATE ALGORITHM=UNDEFINED */
/*!50013 DEFINER=`root`@`127.0.0.1` SQL SECURITY DEFINER */
/*!50001 VIEW `template_properties` AS select `template`.`idtemplate` AS
`idtemplate`,`template`.`name` AS `template`,`template_def`.`idproperty` AS
`idproperty`,`properties`.`name` AS `property`,`properties`.`ptype` AS `property_type`
from ((`template` join `template_def` on((`template`.`idtemplate` =
`template_def`.`idtemplate`))) join `properties` on((`template_def`.`idproperty` =
`properties`.`idproperty`))) */;
/*!50001 SET character_set_client = @saved_cs_client */;
/*!50001 SET character_set_results = @saved_cs_results */;
/*!50001 SET collation_connection = @saved_col_connection */;
--
-- Final view structure for view `values_all`
--
/*!50001 DROP TABLE IF EXISTS `values_all`*/;
/*!50001 DROP VIEW IF EXISTS `values_all`*/;
/*!50001 SET @saved_cs_client = @@character_set_client */;
/*!50001 SET @saved_cs_results = @@character_set_results */;
/*!50001 SET @saved_col_connection = @@collation_connection */;
/*!50001 SET character_set_client = utf8 */;
/*!50001 SET character_set_results = utf8 */;
/*!50001 SET collation_connection = utf8_general_ci */;
/*!50001 CREATE ALGORITHM=UNDEFINED */
/*!50013 DEFINER=`root`@`127.0.0.1` SQL SECURITY DEFINER */
/*!50001 VIEW `values_all` AS select `property_values`.`idstructure` AS
`idstructure`,`properties`.`idproperty` AS `idproperty`,`properties`.`name` AS `name`,NULL
AS `value_string`,`property_values`.`value_num` AS
`value_number`,`properties`.`idreference` AS `idreference` from (`properties` join
`property_values` on((`properties`.`idproperty` = `property_values`.`idproperty`))) where
(`property_values`.`value_num` is not null) union select `property_values`.`idstructure`
AS `idstructure`,`properties`.`idproperty` AS `idproperty`,`properties`.`name` AS
`name`,`property_string`.`value` AS `value_string`,NULL AS
`value_number`,`properties`.`idreference` AS `idreference` from ((`properties` join
`property_values` on((`properties`.`idproperty` = `property_values`.`idproperty`))) join
`property_string` on((`property_values`.`idvalue_string` =
`property_string`.`idvalue_string`))) where (`property_values`.`idvalue_string` is not
null) */;
/*!50001 SET character_set_client = @saved_cs_client */;
/*!50001 SET character_set_results = @saved_cs_results */;
/*!50001 SET collation_connection = @saved_col_connection */;
--
-- Final view structure for view `values_number`
--
/*!50001 DROP TABLE IF EXISTS `values_number`*/;
/*!50001 DROP VIEW IF EXISTS `values_number`*/;
```

```
/*!50001 SET @saved_cs_client = @@character_set_client */;
/*!50001 SET @saved_cs_results = @@character_set_results */;
/*!50001 SET @saved_col_connection = @@collation_connection */;
/*!50001 SET character_set_client = utf8 */;
/*!50001 SET character_set_results = utf8 */;
/*!50001 SET collation_connection = utf8_general_ci */;
/*!50001 CREATE ALGORITHM=UNDEFINED */
/*!50013 DEFINER=`root`@`127.0.0.1` SQL SECURITY DEFINER */
/*!50001 VIEW `values_number` AS select `property_values`.`id` AS
`id`,`property_values`.`idproperty` AS `idproperty`,`property_values`.`idstructure` AS
`idstructure`,`property_values`.`value_num` AS `value`,`property_values`.`status` AS
`status`,`property_values`.`user_name` AS `user_name` from `property_values` where
(`property_values`.`value_num` is not null) */;
/*!50001 SET character_set_client = @saved_cs_client */;
/*!50001 SET character_set_results = @saved_cs_results */;
/*!50001 SET collation_connection = @saved_col_connection */;
--
-- Final view structure for view `values_string`
--
/*!50001 DROP TABLE IF EXISTS `values_string`*/;
/*!50001 DROP VIEW IF EXISTS `values_string`*/;
/*!50001 SET @saved_cs_client = @@character_set_client */;
/*!50001 SET @saved_cs_results = @@character_set_results */;
/*!50001 SET @saved_col_connection = @@collation_connection */;
/*!50001 SET character_set_client = utf8 */;
/*!50001 SET character_set_results = utf8 */;
/*!50001 SET collation_connection = utf8_general_ci */;
/*!50001 CREATE ALGORITHM=UNDEFINED */
/*!50013 DEFINER=`root`@`127.0.0.1` SQL SECURITY DEFINER */
/*!50001 VIEW `values_string` AS select `property_values`.`id` AS
`id`,`properties`.`idproperty` AS `idproperty`,`property_values`.`idstructure` AS
`idstructure`,if((`property_values`.`status` =
'TRUNCATED'),`property_values`.`text`,`property_string`.`value`) AS
`value`,`property_values`.`status` AS `status`,`property_values`.`user_name` AS
`user_name`,`properties`.`name` AS `name` from ((`properties` join `property_values`
on((`properties`.`idproperty` = `property_values`.`idproperty`))) join `property_string`
on((`property_values`.`idvalue_string` = `property_string`.`idvalue_string`))) where
(`property_values`.`idvalue_string` is not null) */;
/*!50001 SET character_set_client = @saved_cs_client */;
/*!50001 SET character_set_results = @saved_cs_results */;
/*!50001 SET collation_connection = @saved_col_connection */;
/*!40103 SET TIME_ZONE=@OLD_TIME_ZONE */;
/*!40101 SET SQL_MODE=@OLD_SQL_MODE */;
/*!40014 SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS */;
/*!40014 SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS */;
/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;
/*!40111 SET SQL_NOTES=@OLD_SQL_NOTES */;
-- Dump completed on 2011-05-30 1:14:43
```

Appendix C: OpenTox Toxicological Endpoints Ontology

www.opentox.org/data/documents/development/RDF%20files/Endpoints/view

A full dump of the OpenTox Toxicological Endpoints Ontology in OWL format is provided below:

```
<?xml version="1.0" encoding="windows-1252"?>
<rdf:RDF
    xmlns:otee="http://www.opentox.org/echaEndpoints.owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_birds">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.12</dc:identifier>
<dc:title>Toxicity to birds</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BAF_fish">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BioAccumulation"/>
</rdfs:subClassOf>
<dc:identifier>2.5.a</dc:identifier>
<dc:title>BAF fish</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Long-
term_toxicity_to_fish_egg_sac_fry_growth_inhibition_of_juvenile_fish_early_life_stage_full_life_cycle
">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.5</dc:identifier>
<dc:title>Long-term toxicity to fish (egg/sac fry, growth inhibition of juvenile fish, early life
stage, full life cycle)</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_earthworms_survival_growth_reproducti
on">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.8</dc:identifier>
<dc:title>Toxicity to earthworms (survival, growth, reproduction)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationAir">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.2</dc:identifier>
<dc:title>Persistence: Abiotic degradation in air (Phototransformation)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Acute_toxicity_to_fish_lethality">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.3</dc:identifier>
<dc:title>Acute toxicity to fish (lethality)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Ready_not_ready_biodegradability">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#PersistenceBiodegradation"/>
</rdfs:subClassOf>
<dc:identifier>2.3.a</dc:identifier>
<dc:title>Ready/not ready biodegradability</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Long-
term_toxicity_survival_growth_reproduction">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_birds"/>
<dc:identifier>3.12.b</dc:identifier>
<dc:title>Long-term toxicity (survival, growth, reproduction)</dc:title>
```

```
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Endpoints"/>
</rdfs:subClassOf>
<dc:identifier>1</dc:identifier>
<dc:title>Physicochemical effects </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Endpoints"/>
</rdfs:subClassOf>
<dc:identifier>5</dc:identifier>
<dc:title>Toxicokinetics </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#AcuteInhalationToxicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.1</dc:identifier>
<dc:title>Acute inhalation toicity</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Other_e.g._inhibition_of_specific_enzymes_involve
d_in_hormone_synthesis_or_regulation_specify_enzyme_s_and_hormone">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EndocrineActivity"/>
</rdfs:subClassOf>
<dc:identifier>4.18.c</dc:identifier>
<dc:title>Other (e.g. inhibition of specific enzymes involved in hormone synthesis or regulation,
specify enzyme(s) and hormone)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Blood-brain_barrier_penetration">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.4</dc:identifier>
<dc:title>Blood-brain barrier penetration</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Protein-binding">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.9</dc:identifier>
<dc:title>Protein-binding</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Vegetation-
soil_partition_coefficient">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.11</dc:identifier>
<dc:title>Vegetation-soil partition coefficient</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Biodegradation_time_frame_primary_ultimate_degrad
ation">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#PersistenceBiodegradation"/>
</rdfs:subClassOf>
<dc:identifier>2.3.b</dc:identifier>
<dc:title>Biodegradation time frame (primary, ultimate degradation)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BAF_other_organisms">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BioAccumulation"/>
</rdfs:subClassOf>
<dc:identifier>2.5.b</dc:identifier>
<dc:title>BAF other organisms </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Direct_photolysis">
<rdfs:subClassOf
rdf:resource="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationAir"/>
<dc:identifier>2.2.a</dc:identifier>
<dc:title>Direct photolysis</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Short-
term_toxicity_to_Daphnia_immobilisation">
```

```
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.1</dc:identifier>
<dc:title>Short-term toxicity to Daphnia (immobilisation)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Short-
term_toxicity_to_algae_inhibition_of_the_exponential_growth_rate">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.2</dc:identifier>
<dc:title>Short-term toxicity to algae (inhibition of the exponential growth rate)</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_soil_invertebrates_survival_growth_re
production">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
</rdfs:subClassOf>
<dc:identifier>3.10</dc:identifier>
<dc:title>Toxicity to soil invertebrates (survival, growth, reproduction)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Adsorption_Desorption_in_soil">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.7</dc:identifier>
<dc:title>Adsorption/Desorption in soil </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Endpoints"/>
</rdfs:subClassOf>
<dc:identifier>3</dc:identifier>
<dc:title>Ecotoxic effects</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Eye_irritation_corrosion">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.9</dc:identifier>
<dc:title>Eye irritation/corrosion</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Organic_carbon-
sorption_partition_coefficient_organic_carbon_Koc">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.6</dc:identifier>
<dc:title>Organic carbon-sorption partition coefficient (organic carbon; Koc)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Indirect_photolysis_OH-
radical_reaction_ozone-radical_reaction_other">
<rdfs:subClassOf
rdf:resource="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationAir"/>
<dc:identifier>2.2.b</dc:identifier>
<dc:title>Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Mutagenicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.10</dc:identifier>
<dc:title>Mutagenicity </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Placental_barrier_penetration">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.5</dc:identifier>
<dc:title>Placental barrier penetration</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Dissociation_constant_pKa">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.10</dc:identifier>
```

<dc:title>Dissociation constant (pKa)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#PersistenceBiodegradation">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.3</dc:identifier>
<dc:title>Persistence: Biodegradation</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Adsorption_Desorption_in_sediment">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.8</dc:identifier>
<dc:title>Adsorption/Desorption in sediment</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#In_vivo_pre-natal-
developmental_toxicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.16</dc:identifier>
<dc:title>In vivo pre-natal-developmental toxicity</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Blood-testis_barrier_penetration">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.6</dc:identifier>
<dc:title>Blood-testis barrier penetration</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Skin_penetration">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.1</dc:identifier>
<dc:title>Skin penetration</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Surface_tension">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.5</dc:identifier>
<dc:title>Surface tension</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#AcuteOralToxicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.2</dc:identifier>
<dc:title>Acute oral toxicity</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EndocrineActivity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.18</dc:identifier>
<dc:title>Endocrine Activity</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Acute_photoirritation">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.5</dc:identifier>
<dc:title>Acute photoirritation</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationWater">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.1</dc:identifier>
<dc:title>Persistence: Abiotic degradation in water</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#In_vitro_reproductive_toxicity_e.g._embryotoxic_e
ffects_in_cell_culture_such_as_embryo_stem_cells">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.15</dc:identifier>

<dc:title>In vitro reproductive toxicity (e.g. embryotoxic effects in cell culture such as embryo stem cells) </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#SkinIrritationCorrosion">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.4</dc:identifier>
<dc:title>Skin irritation /corrosion</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_plants_leaves_seed_germination_root_elongation">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
<dc:identifier>3.9</dc:identifier>
<dc:title>Toxicity to plants (leaves, seed germination, root elongation)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#SkinSensitisation">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.6</dc:identifier>
<dc:title>Skin sensitisation</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#In_vivo_pre-_peri-_post_natal_development_and_or_fertility_1_or_2_gen._Study_or_enhanced_1_gen_study">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.17</dc:identifier>
<dc:title>In vivo pre-, peri-, post natal development and / or fertility (1 or 2 gen. Study or enhanced 1 gen study) </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Gastrointestinal_absorption">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.3</dc:identifier>
<dc:title>Gastrointestinal absorption</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Water_solubility">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.3</dc:identifier>
<dc:title>Water solubility</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Octanol-water_partition_coefficient_Kow">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.6</dc:identifier>
<dc:title>Octanol-water partition coefficient (Kow)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Air-_water_partition_coefficient_Henry_s_law_constant_H">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.9</dc:identifier>
<dc:title>Air- water partition coefficient (Henry`s law constant, H)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Hydrolysis">
<rdfs:subClassOf
rdf:resource="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationWater"/>
<dc:identifier>2.1.a</dc:identifier>
<dc:title>Hydrolysis </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BCF_other_organisms">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BioConcentration"/>
</rdfs:subClassOf>
<dc:identifier>2.5.b</dc:identifier>
<dc:title>BCF other organisms </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Receptor-binding">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EndocrineActivity"/>
<dc:identifier>4.18.a</dc:identifier>
<dc:title>Receptor-binding (specify receptor)</dc:title>
</owl:Class>

```
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Vegetation-
water_partition_coefficient">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.9</dc:identifier>
<dc:title>Vegetation-water partition coefficient</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BioAccumulation">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.5</dc:identifier>
<dc:title>Bioaccumulation</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Photomutagenicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.11</dc:identifier>
<dc:title>Photomutagenicity</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BCF_fish">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BioConcentration"/>
</rdfs:subClassOf>
<dc:identifier>2.4.a</dc:identifier>
<dc:title>BCF fish</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Respiratory_sensitisation">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.7</dc:identifier>
<dc:title>Respiratory sensitisation</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Oxidation">
<rdfs:subClassOf
rdf:resource="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationWater"/>
<dc:identifier>2.1.b</dc:identifier>
<dc:title>Oxidation </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Blood-lung_barrier_penetration">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.7</dc:identifier>
<dc:title>Blood-lung barrier penetration</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Vegetation-air_partition_coefficient">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
</rdfs:subClassOf>
<dc:identifier>2.10</dc:identifier>
<dc:title>Vegetation-air partition coefficient</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Vapour_pressure">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.4</dc:identifier>
<dc:title>Vapour pressure</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Persistence_Others">
<rdfs:subClassOf
rdf:resource="http://www.opentox.org/echaEndpoints.owl#PersistenceAbioticDegradationWater"/>
<dc:identifier>2.1.c</dc:identifier>
<dc:title>Others</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Endpoints"/>
</rdfs:subClassOf>
<dc:identifier>2</dc:identifier>
<dc:title>Environmental fate parameters </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#OcularMembranePenetration">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
```

```
<dc:identifier>5.2</dc:identifier>
<dc:title>Ocular membrane penetration</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Long-
term_toxicity_to_Daphnia_lethality_inhibition_of_reproduction">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
<dc:identifier>3.4</dc:identifier>
<dc:title>Long-term toxicity to Daphnia (lethality, inhibition of reproduction)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Endpoints">
<dc:title>Endpoints</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Octanol-
air_partition_coefficient_Koa">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.8</dc:identifier>
<dc:title>Octanol-air partition coefficient (Koa)</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_sediment_organisms_survival_growth_re
production">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
<dc:identifier>3.11</dc:identifier>
<dc:title>Toxicity to sediment organisms (survival, growth, reproduction)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Receptor_binding_and_gene_expression">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EndocrineActivity"/>
<dc:identifier>4.18.b</dc:identifier>
<dc:title>Receptor binding and gene expression (specify receptor)</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Microbial_inhibition_activated_sludge_respiration
_inhibition_inhibition_of_nitrification_other">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
<dc:identifier>3.6</dc:identifier>
<dc:title>Microbial inhibition (activated sludge respiration inhibition, inhibition of nitrification,
other)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#AcuteDermalToxicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.3</dc:identifier>
<dc:title>Acute dermal toxicity</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#BioConcentration">
<rdfs:subClassOf
rdf:resource="http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters"/>
<dc:identifier>2.4</dc:identifier>
<dc:title>Bioconcentration </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Octanol-
water_distribution_coefficient_D">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.7</dc:identifier>
<dc:title>Octanol-water distribution coefficient (D)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Metabolism">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.8</dc:identifier>
<dc:title>Metabolism (including metabolic clearance)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Melting_point">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.1</dc:identifier>
<dc:title>Melting point</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Photocarcinogenicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.13</dc:identifier>
<dc:title>Photocarcinogenicity</dc:title>
</owl:Class>
```

<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Repeated_dose_toxicity">
<rdfs:subClassOf>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
</rdfs:subClassOf>
<dc:identifier>4.14</dc:identifier>
<dc:title>Repeated dose toxicity </dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#Endpoints"/>
<dc:identifier>4</dc:identifier>
<dc:title>Human health effects</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Photosensitisation">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
<dc:identifier>4.8</dc:identifier>
<dc:title>Photosensitisation</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_soil_microorganisms_inhibition_of_C-mineralisation_inhibition_of_N-mineralisation_other">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects"/>
<dc:identifier>3.7</dc:identifier>
<dc:title>Toxicity to soil microorganisms (inhibition of C-mineralisation, inhibition of N-mineralisation, other)</dc:title>
</owl:Class>
<owl:Class
rdf:about="http://www.opentox.org/echaEndpoints.owl#Short_term_toxicity_feeding_gavage_other">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#Toxicity_to_birds"/>
<dc:identifier>3.12.a</dc:identifier>
<dc:title>Short term toxicity (feeding, gavage, other)</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Boiling_point">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects"/>
<dc:identifier>1.2</dc:identifier>
<dc:title>Boiling point</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#Carcinogenicity">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects"/>
<dc:identifier>4.12</dc:identifier>
<dc:title>Carcinogenicity</dc:title>
</owl:Class>
<owl:Class rdf:about="http://www.opentox.org/echaEndpoints.owl#DNA-binding">
<rdfs:subClassOf rdf:resource="http://www.opentox.org/echaEndpoints.owl#ToxicoKinetics"/>
<dc:identifier>5.10</dc:identifier>
<dc:title>DNA-binding</dc:title>
</owl:Class>
<otee:SkinSensitisation rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Skin sensitisation">
<dc:title>Skin sensitisation</dc:title>
</otee:SkinSensitisation>
<otee:Other_e.g._inhibition_of_specific_enzymes_involved_in_hormone_synthesis_or_regulation_specify_enzyme_s_and_hormone rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Other (e_g_ inhibition of specific enzymes involved in hormone synthesis or regulation, specify enzyme(s) and hormone)">
<dc:title>Other (e.g. inhibition of specific enzymes involved in hormone synthesis or regulation, specify enzyme(s) and hormone)</dc:title>
</otee:Other_e.g._inhibition_of_specific_enzymes_involved_in_hormone_synthesis_or_regulation_specify_enzyme_s_and_hormone>
<otee:BCF_other_organisms rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_BCF other organisms">
<dc:title>BCF other organisms </dc:title>
</otee:BCF_other_organisms>
<otee:Protein-binding rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Protein-binding">
<dc:title>Protein-binding</dc:title>
</otee:Protein-binding>
<otee:BioAccumulation rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Bioaccumulation">
<dc:title>Bioaccumulation</dc:title>
</otee:BioAccumulation>
<otee:In_vitro_reproductive_toxicity_e.g._embryotoxic_effects_in_cell_culture_such_as_embryo_stem_cells rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_In vitro reproductive toxicity (e_g_ embryotoxic effects in cell culture such as embryo stem cells)">
<dc:title>In vitro reproductive toxicity (e.g. embryotoxic effects in cell culture such as embryo stem cells) </dc:title>

```
</otee:In_vitro_reproductive_toxicity_e.g._embryotoxic_effects_in_cell_culture_such_as_embryo_stem_ce
lls>
<otee:PersistenceBiodegradation
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Persistence: Biodegradation">
<dc:title>Persistence: Biodegradation</dc:title>
</otee:PersistenceBiodegradation>
<otee:Toxicity_to_birds rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicity to
birds">
<dc:title>Toxicity to birds</dc:title>
</otee:Toxicity_to_birds>
<otee:Vapour_pressure rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Vapour pressure">
<dc:title>Vapour pressure</dc:title>
</otee:Vapour_pressure>
<otee:Ready_not_ready_biodegradability
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Ready/not ready biodegradability">
<dc:title>Ready/not ready biodegradability</dc:title>
</otee:Ready_not_ready_biodegradability>
<otee:Blood-brain_barrier_penetration
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Blood-brain barrier penetration">
<dc:title>Blood-brain barrier penetration</dc:title>
</otee:Blood-brain_barrier_penetration>
<otee:Receptor_binding_and_gene_expression
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Receptor binding and gene expression
(specify receptor)">
<dc:title>Receptor binding and gene expression (specify receptor)</dc:title>
</otee:Receptor_binding_and_gene_expression>
<otee:Skin_penetration rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Skin
penetration">
<dc:title>Skin penetration</dc:title>
</otee:Skin_penetration>
<otee:OcularMembranePenetration rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Ocular
membrane penetration">
<dc:title>Ocular membrane penetration</dc:title>
</otee:OcularMembranePenetration>
<otee:EndocrineActivity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Endocrine
Activity">
<dc:title>Endocrine Activity</dc:title>
</otee:EndocrineActivity>
<otee:DNA-binding rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_DNA-binding">
<dc:title>DNA-binding</dc:title>
</otee:DNA-binding>
<otee:Gastrointestinal_absorption
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Gastrointestinal absorption">
<dc:title>Gastrointestinal absorption</dc:title>
</otee:Gastrointestinal_absorption>
<otee:AcuteDermalToxicity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Acute dermal
toxicity">
<dc:title>Acute dermal toxicity</dc:title>
</otee:AcuteDermalToxicity>
<otee:Blood-lung_barrier_penetration
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Blood-lung barrier penetration">
<dc:title>Blood-lung barrier penetration</dc:title>
</otee:Blood-lung_barrier_penetration>
<otee:Dissociation_constant_pKa
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Dissociation constant (pKa)">
<dc:title>Dissociation constant (pKa)</dc:title>
</otee:Dissociation_constant_pKa>
<otee:Long-term_toxicity_to_Daphnia_lethality_inhibition_of_reproduction
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Long-term toxicity to Daphnia
(lethality, inhibition of reproduction)">
<dc:title>Long-term toxicity to Daphnia (lethality, inhibition of reproduction)</dc:title>
</otee:Long-term_toxicity_to_Daphnia_lethality_inhibition_of_reproduction>
<otee:Vegetation-soil_partition_coefficient
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Vegetation-soil partition coefficient">
<dc:title>Vegetation-soil partition coefficient</dc:title>
</otee:Vegetation-soil_partition_coefficient>
<otee:SkinIrritationCorrosion rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Skin
irritation /corrosion">
<dc:title>Skin irritation /corrosion</dc:title>
</otee:SkinIrritationCorrosion>
<otee:Adsorption_Desorption_in_soil
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Adsorption/Desorption in soil">
<dc:title>Adsorption/Desorption in soil </dc:title>
```

```
</otee:Adsorption_Desorption_in_soil>
<otee:AcuteOralToxicity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Acute oral
toxicity">
<dc:title>Acute oral toxicity</dc:title>
</otee:AcuteOralToxicity>
<otee:Toxicity_to_sediment_organisms_survival_growth_reproduction
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicity to sediment organisms
(survival, growth, reproduction)">
<dc:title>Toxicity to sediment organisms (survival, growth, reproduction)</dc:title>
</otee:Toxicity_to_sediment_organisms_survival_growth_reproduction>
<otee:Photosensitisation
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Photosensitisation">
<dc:title>Photosensitisation</dc:title>
</otee:Photosensitisation>
<otee:Water_solubility rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Water
solubility">
<dc:title>Water solubility</dc:title>
</otee:Water_solubility>
<otee:BioConcentration
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Bioconcentration">
<dc:title>Bioconcentration </dc:title>
</otee:BioConcentration>
<otee:Acute_photoirritation rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Acute
photoirritation">
<dc:title>Acute photoirritation</dc:title>
</otee:Acute_photoirritation>
<otee:Short-term_toxicity_to_Daphnia_immobilisation
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Short-term toxicity to Daphnia
(immobilisation)">
<dc:title>Short-term toxicity to Daphnia (immobilisation)</dc:title>
</otee:Short-term_toxicity_to_Daphnia_immobilisation>
<otee:Short_term_toxicity_feeding_gavage_other
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Short term toxicity (feeding, gavage,
other)">
<dc:title>Short term toxicity (feeding, gavage, other)</dc:title>
</otee:Short_term_toxicity_feeding_gavage_other>
<otee:Respiratory_sensitisation
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Respiratory sensitisation">
<dc:title>Respiratory sensitisation</dc:title>
</otee:Respiratory_sensitisation>
<otee:Air-_water_partition_coefficient_Henry_s_law_constant_H
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Air- water partition coefficient
(Henry`s law constant, H)">
<dc:title>Air- water partition coefficient (Henry`s law constant, H)</dc:title>
</otee:Air-_water_partition_coefficient_Henry_s_law_constant_H>
<otee:Blood-testis_barrier_penetration
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Blood-testis barrier penetration">
<dc:title>Blood-testis barrier penetration</dc:title>
</otee:Blood-testis_barrier_penetration>
<otee:PersistenceAbioticDegradationAir
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Persistence: Abiotic degradation in air
(Phototransformation)">
<dc:title>Persistence: Abiotic degradation in air (Phototransformation)</dc:title>
</otee:PersistenceAbioticDegradationAir>
<owl:AnnotationProperty rdf:about="http://purl.org/dc/elements/1.1/title"/>
<otee:ToxicoKinetics rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicokinetics">
<dc:title>Toxicokinetics </dc:title>
</otee:ToxicoKinetics>
<otee:Eye_irritation_corrosion rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Eye
irritation/corrosion">
<dc:title>Eye irritation/corrosion</dc:title>
</otee:Eye_irritation_corrosion>
<owl:AnnotationProperty rdf:about="http://purl.org/dc/elements/1.1/identifier"/>
<otee:Acute_toxicity_to_fish_lethality
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Acute toxicity to fish (lethality)">
<dc:title>Acute toxicity to fish (lethality)</dc:title>
</otee:Acute_toxicity_to_fish_lethality>
<otee:Receptor-binding rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Receptor-binding
(specify receptor)">
<dc:title>Receptor-binding (specify receptor)</dc:title>
</otee:Receptor-binding>
<otee:Placental_barrier_penetration
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Placental barrier penetration">
<dc:title>Placental barrier penetration</dc:title>
```

```
</otee:Placental_barrier_penetration>
<otee:Direct_photolysis rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Direct
photolysis">
<dc:title>Direct photolysis</dc:title>
</otee:Direct_photolysis>
<otee:In_vivo_pre-natal-developmental_toxicity
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_In vivo pre-natal-developmental
toxicity">
<dc:title>In vivo pre-natal-developmental toxicity</dc:title>
</otee:In_vivo_pre-natal-developmental_toxicity>
<otee:PhysicoChemicalEffects
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Physicochemical effects">
<dc:title>Physicochemical effects </dc:title>
</otee:PhysicoChemicalEffects>
<otee:Mutagenicity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Mutagenicity">
<dc:title>Mutagenicity </dc:title>
</otee:Mutagenicity>
<otee:Adsorption_Desorption_in_sediment
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Adsorption/Desorption in sediment">
<dc:title>Adsorption/Desorption in sediment</dc:title>
</otee:Adsorption_Desorption_in_sediment>
<otee:Short-term_toxicity_to_algae_inhibition_of_the_exponential_growth_rate
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Short-term toxicity to algae (inhibition
of the exponential growth rate)">
<dc:title>Short-term toxicity to algae (inhibition of the exponential growth rate)</dc:title>
</otee:Short-term_toxicity_to_algae_inhibition_of_the_exponential_growth_rate>
<otee:Surface_tension rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Surface tension">
<dc:title>Surface tension</dc:title>
</otee:Surface_tension>
<otee:Long-term_toxicity_survival_growth_reproduction
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Long-term toxicity (survival, growth,
reproduction)">
<dc:title>Long-term toxicity (survival, growth, reproduction)</dc:title>
</otee:Long-term_toxicity_survival_growth_reproduction>
<otee:Toxicity_to_plants_leaves_seed_germination_root_elongation
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicity to plants (leaves, seed
germination, root elongation)">
<dc:title>Toxicity to plants (leaves, seed germination, root elongation)</dc:title>
</otee:Toxicity_to_plants_leaves_seed_germination_root_elongation>
<otee:Organic_carbon-sorption_partition_coefficient_organic_carbon_Koc
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Organic carbon-sorption partition
coefficient (organic carbon; Koc)">
<dc:title>Organic carbon-sorption partition coefficient (organic carbon; Koc)</dc:title>
</otee:Organic_carbon-sorption_partition_coefficient_organic_carbon_Koc>
<otee:Oxidation rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Oxidation">
<dc:title>Oxidation </dc:title>
</otee:Oxidation>
<otee:Vegetation-water_partition_coefficient
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Vegetation-water partition coefficient">
<dc:title>Vegetation-water partition coefficient</dc:title>
</otee:Vegetation-water_partition_coefficient>
<otee:HumanHealthEffects rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Human health
effects">
<dc:title>Human health effects</dc:title>
</otee:HumanHealthEffects>
<otee:Melting_point rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Melting point">
<dc:title>Melting point</dc:title>
</otee:Melting_point>
<otee:Persistence_Others rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Others">
<dc:title>Others</dc:title>
</otee:Persistence_Others>
<otee:BCF_fish rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_BCF fish">
<dc:title>BCF fish</dc:title>
</otee:BCF_fish>
<otee:Hydrolysis rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Hydrolysis">
<dc:title>Hydrolysis </dc:title>
</otee:Hydrolysis>
<otee:Vegetation-air_partition_coefficient
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Vegetation-air partition coefficient">
<dc:title>Vegetation-air partition coefficient</dc:title>
</otee:Vegetation-air_partition_coefficient>
<otee:Octanol-water_distribution_coefficient_D
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Octanol-water distribution coefficient
(D)">
```

<dc:title>Octanol-water distribution coefficient (D)</dc:title>
</otee:Octanol-water_distribution_coefficient_D>
<otee:In_vivo_pre-_peri-
_post_natal_development_and_or_fertility_1_or_2_gen._Study_or_enhanced_1_gen_study
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_In vivo pre-, peri-, post natal
development and / or fertility (1 or 2 gen_ Study or enhanced 1 gen study)">
<dc:title>In vivo pre-, peri-, post natal development and / or fertility (1 or 2 gen. Study or
enhanced 1 gen study) </dc:title>
</otee:In_vivo_pre-_peri-
_post_natal_development_and_or_fertility_1_or_2_gen._Study_or_enhanced_1_gen_study>
<otee:Carcinogenicity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Carcinogenicity">
<dc:title>Carcinogenicity</dc:title>
</otee:Carcinogenicity>
<otee:PersistenceAbioticDegradationWater
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Persistence: Abiotic degradation in
water">
<dc:title>Persistence: Abiotic degradation in water</dc:title>
</otee:PersistenceAbioticDegradationWater>
<otee:AcuteInhalationToxicity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Acute
inhalation toicity">
<dc:title>Acute inhalation toicity</dc:title>
</otee:AcuteInhalationToxicity>
<otee:BAF_other_organisms rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_BAF other
organisms">
<dc:title>BAF other organisms </dc:title>
</otee:BAF_other_organisms>
<otee:Repeated_dose_toxicity rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Repeated
dose toxicity">
<dc:title>Repeated dose toxicity </dc:title>
</otee:Repeated_dose_toxicity>
<otee:Octanol-air_partition_coefficient_Koa
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Octanol-air partition coefficient
(Koa)">
<dc:title>Octanol-air partition coefficient (Koa)</dc:title>
</otee:Octanol-air_partition_coefficient_Koa>
<otee:Toxicity_to_soil_microorganisms_inhibition_of_C-mineralisation_inhibition_of_N-
mineralisation_other rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicity to soil
microorganisms (inhibition of C-mineralisation, inhibition of N-mineralisation, other)">
<dc:title>Toxicity to soil microorganisms (inhibition of C-mineralisation, inhibition of N-
mineralisation, other)</dc:title>
</otee:Toxicity_to_soil_microorganisms_inhibition_of_C-mineralisation_inhibition_of_N-
mineralisation_other>
<otee:Indirect_photolysis_OH-radical_reaction_ozone-radical_reaction_other
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Indirect photolysis (OH-radical
reaction, ozone-radical reaction, other)">
<dc:title>Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)</dc:title>
</otee:Indirect_photolysis_OH-radical_reaction_ozone-radical_reaction_other>
<otee:BAF_fish rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_BAF fish">
<dc:title>BAF fish</dc:title>
</otee:BAF_fish>
<otee:Octanol-water_partition_coefficient_Kow
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Octanol-water partition coefficient
(Kow)">
<dc:title>Octanol-water partition coefficient (Kow)</dc:title>
</otee:Octanol-water_partition_coefficient_Kow>
<otee:Photocarcinogenicity
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Photocarcinogenicity">
<dc:title>Photocarcinogenicity</dc:title>
</otee:Photocarcinogenicity>
<otee:Toxicity_to_soil_invertebrates_survival_growth_reproduction
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicity to soil invertebrates
(survival, growth, reproduction)">
<dc:title>Toxicity to soil invertebrates (survival, growth, reproduction)</dc:title>
</otee:Toxicity_to_soil_invertebrates_survival_growth_reproduction>
<otee:EcotoxicEffects rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Ecotoxic effects">
<dc:title>Ecotoxic effects</dc:title>
</otee:EcotoxicEffects>
<otee:Boiling_point rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Boiling point">
<dc:title>Boiling point</dc:title>
</otee:Boiling_point>
<otee:EnvironmentalFateParameters
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Environmental fate parameters">
<dc:title>Environmental fate parameters </dc:title>
</otee:EnvironmentalFateParameters>

<otee:Long-
term_toxicity_to_fish_egg_sac_fry_growth_inhibition_of_juvenile_fish_early_life_stage_full_life_cycle
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Long-term toxicity to fish (egg/sac fry,
growth inhibition of juvenile fish, early life stage, full life cycle)">
<dc:title>Long-term toxicity to fish (egg/sac fry, growth inhibition of juvenile fish, early life
stage, full life cycle)</dc:title>
</otee:Long-
term_toxicity_to_fish_egg_sac_fry_growth_inhibition_of_juvenile_fish_early_life_stage_full_life_cycle
>
<otee:Toxicity_to_earthworms_survival_growth_reproduction
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Toxicity to earthworms (survival,
growth, reproduction)">
<dc:title>Toxicity to earthworms (survival, growth, reproduction)</dc:title>
</otee:Toxicity_to_earthworms_survival_growth_reproduction>
<otee:Metabolism rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Metabolism (including
metabolic clearance)">
<dc:title>Metabolism (including metabolic clearance)</dc:title>
</otee:Metabolism>
<otee:Microbial_inhibition_activated_sludge_respiration_inhibition_inhibition_of_nitrification_other
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Microbial inhibition (activated sludge
respiration inhibition, inhibition of nitrification, other)">
<dc:title>Microbial inhibition (activated sludge respiration inhibition, inhibition of nitrification,
other)</dc:title>
</otee:Microbial_inhibition_activated_sludge_respiration_inhibition_inhibition_of_nitrification_other
>
<otee:Photomutagenicity
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Photomutagenicity">
<dc:title>Photomutagenicity</dc:title>
</otee:Photomutagenicity>
<otee:Biodegradation_time_frame_primary_ultimate_degradation
rdf:about="http://www.opentox.org/echaEndpoints.owl#ENDPOINT_Biodegradation time frame (primary,
ultimate degradation)">
<dc:title>Biodegradation time frame (primary, ultimate degradation)</dc:title>
</otee:Biodegradation_time_frame_primary_ultimate_degradation>
</rdf:RDF>