Deliverable D3.2

Prototype Database

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 / D3.2	
Name	Prototype Database	
Document ID:	OpenTox Deliverable Report WP3-D3.2	
Date:	Feb 27, 2010	
Status:	Final Version	
Organisation:	Ideaconsult Ltd. (IDEA)	
Contributors	Vedrin Jeliazkov	IDEA
	Nina Jeliazkova	IDEA
	Pantelis Sopasakis	NTUA
	Alexey Lagunin	IBMC
	Romualdo Benigni	ISS
	Barry Hardy (review)	DC
	Stefan Kramer (review)	TUM

Distribution:	Public

Purpose of Document:	To document results for this deliverable

Document History:	a – First draft, circulated on Feb 15, 2010
	b – Second draft, circulated on Feb 17, 2010
	c – Third draft, circulated on Feb 23, 2010
	d – Fourth draft, circulated on Feb 24, 2010
	e – Fifth draft, circulated on Feb 26, 2010
	f – Sixth draft, circulated on Feb 27, 2010
	g – Final Version, approved Feb 27, 2010
	H – Updated Version, Mar 8, 2010

Table of Contents

Table of Figures

List of Tables

Summary

A major pre-requisite 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 chemical properties. In particular, the range of replacement alternatives 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 it is likely that, in many circumstances, an animal test cannot be currently replaced by a single replacement alternative method, the development, evaluation and optimisation of stepwise testing strategies and integrated testing schemes should be encouraged. The OpenTox data facilities, made publically accessible through a web services framework, provide a solid basis for addressing the above mentioned replacement alternative goals in a more efficient, technically sound and integrated way compared to current uncoordinated practices and fragmented resources. Unfortunately, even today, more than half a century after Russell and Burch's original publication and 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 or even 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:

- 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;
- 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, through the OpenTox Framework.

The OpenTox prototype database includes ECHA's list of pre-registered substances[1] along with high-quality data from consortium members (e.g. ISS ISSCAN[2], IDEA AMBIT[3]) and third parties (e.g. JRC PRS list[4], EPA

[1] apps.echa.europa.eu/preregistered/pre-registered-sub.aspx accessed on Feb 14, 2010

[2] www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7 accessed on Feb 14, 2010

DSSTox[5], ECETOC skin irritation[6], LLNA skin sensitisation[7], Bioconcentration factor (BCF) Gold Standard Database[8]). Additional data for chemical structures has been collected from various public sources (e.g. Chemical Identifier Resolver[9], ChemIDplus[10], PubChem[11]) and further checked manually by experts. The database provides means to identify the origin of the data, i.e., the specific inventory a compound originated from. The data is currently publicly available and accessible via initial implementation of REST web services[12], as defined in the OpenTox Framework design and its implementations. We also describe an OWL (Web Ontology Language)[13] ontology of toxicological endpoints, which corresponds to the endpoint classification of REACH guidance documents[14] and allows for unique mapping between endpoints from various inventories.

This report describes in detail the above-mentioned OpenTox data facilities and resources, as well as future directions and extensions.

---

[3] ambit.sourceforge.net accessed on Feb 14, 2010

[4] ecb.jrc.ec.europa.eu/home.php?CONTENU=/DOCUMENTS/QSAR/INFORMATION_SOURCES/EC_CHEMICAL_INVENTORIES accessed on Feb 14, 2010

[5] www.epa.gov/ncct/dsstox/ accessed on Feb 14, 2010

[6] ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)

[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.

[8] ambit.sourceforge.net/euras accessed on Feb 14, 2010

[9] cactus.nci.nih.gov/chemical/structure accessed on Feb 14, 2010

[10] chem.sis.nlm.nih.gov/chemidplus accessed on Feb 14, 2010

[11] pubchem.ncbi.nlm.nih.gov accessed on Feb 14, 2010

[12] http://apps.ideaconsult.net:8180/ambit2/ accessed on Feb 14, 2010

[13] www.w3.org/TR/owl-features accessed on Feb 14, 2010

[14] guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?%20vers=20_08_08 accessed on Feb 14, 2010

1. AMBIT Software and Database

1.1 Overview

Ambit is an open source software for cheminformatics data management. It consists of a database and functional modules allowing for a variety of queries and data mining of the information, stored in the database. Ambit XT is a standalone application, with a graphical user interface, based on Ambit modules. Ambit functionality is being extended and wrapped as REST[15] web services within the framework of the OpenTox project. Ambit (source code, executables and a link to REST services) is available at ambit.sourceforge.net.

Ambit software is organised in several modules with well defined dependency, as highlighted in Table 1.

Table 1 Ambit software modules

Module	Description
AmbitXT	GUI application
AmbitXT plugin: Database search and Analogue identification	AmbitXT plugin, allowing various database queries and analogues identification
AmbitXT plugin: Category building	AmbitXT plugin for analogues identification
AmbitXT plugin: Database tools	AmbitXT plugin for database import and management
AmbitXT plugin: Database administration	AmbitXT plugin for database administration activities
AmbitXT plugin: REACH PBT assessment	AmbitXT plugin, implementing a workflow for REACH-compliant Persistence, Biodegradation and Toxicity (PBT) assessment
ambit2-base	Base classes, without cheminformatics functionality
ambit2-core	Core classes, with cheminformatics functionality
ambit2-hashcode	Hash codes
ambit2-smarts	SMARTS parser
ambit2-db	Database functionality
ambit2-smi23d	Wrapper for Smi23d executables cicc-grid.svn.sourceforge.net/viewvc/cicc-grid/cicc-grid/smi23d/trunk
ambit2-mopac	Wrapper for OpenMopac sourceforge.net/projects/mopac7
ambit2-ui	User interface
ambit2-dbui	Database user interface
ambit2-workflow	Workflow module

[15] www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm accessed on Feb 14, 2010

Module	Description
ambit2-namestructure	Chemical name to structure convertor, based on OPSIN package sourceforge.net/projects/oscar3-chem/files
ambit2-model	Similarity calculation, feature selection and QSAR model development
ambit2-taglibs	JSP tags
PubChem utilities	PubChem access utilities
Ambit2 REST web services	Allows querying Ambit database through REST style web services

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[16]. A very brief summary of the tables in the Ambit database is provided in Table 2.

Table 2 Ambit database tables summary

Table	Description
Chemical structures	
chemicals	Chemical compounds
structure	Chemical structures, conformers
history	Previous versions of chemical structures
Inventories	
src_dataset	Datasets
struc_dataset	Lookup table for structures, belonging to a dataset
Identifiers, Descriptors, Properties	
catalog_references	References
properties	Property definition (name, reference, units)
property_values	Numerical property values or links to string values
property_string	String values
property_tuples	Tuples of properties
tuples	Tuples per dataset
template	Templates

[16] www.mysql.com accessed on Feb 14, 2010

Table	Description
template_def	Template definition (which properties belong to a template)
dictionary	Templates hierarchy
Queries	
query	Queries
query_results	Structures per query
sessions	Sessions
Users support	
user_roles	Roles, assigned to users
roles	User roles
users	Users
Models	
(Q)SAR models	A holder for a model, including pointers to the training dataset (query table), set of predictors (template table) and set of dependent variables (template table)
Quality assessment support	
quality_chemicals	Quality labels of structures and properties
quality_labels	
quality_pair	
quality_structure	
Pre-processed data for substructure, similarity and SMARTS queries	
fp1024	Pre-processed fingerprints for pre-screening and similarity search
fp1024_struc	
sk1024	Pre-processed fragments for accelerating SMARTS searches
atom_distance	Pre-processed data for atom environments similarity
atom_structure	
Schema version	
version	Database version

An overview of the entity-relationship diagram of the database is provided in Figure 1.

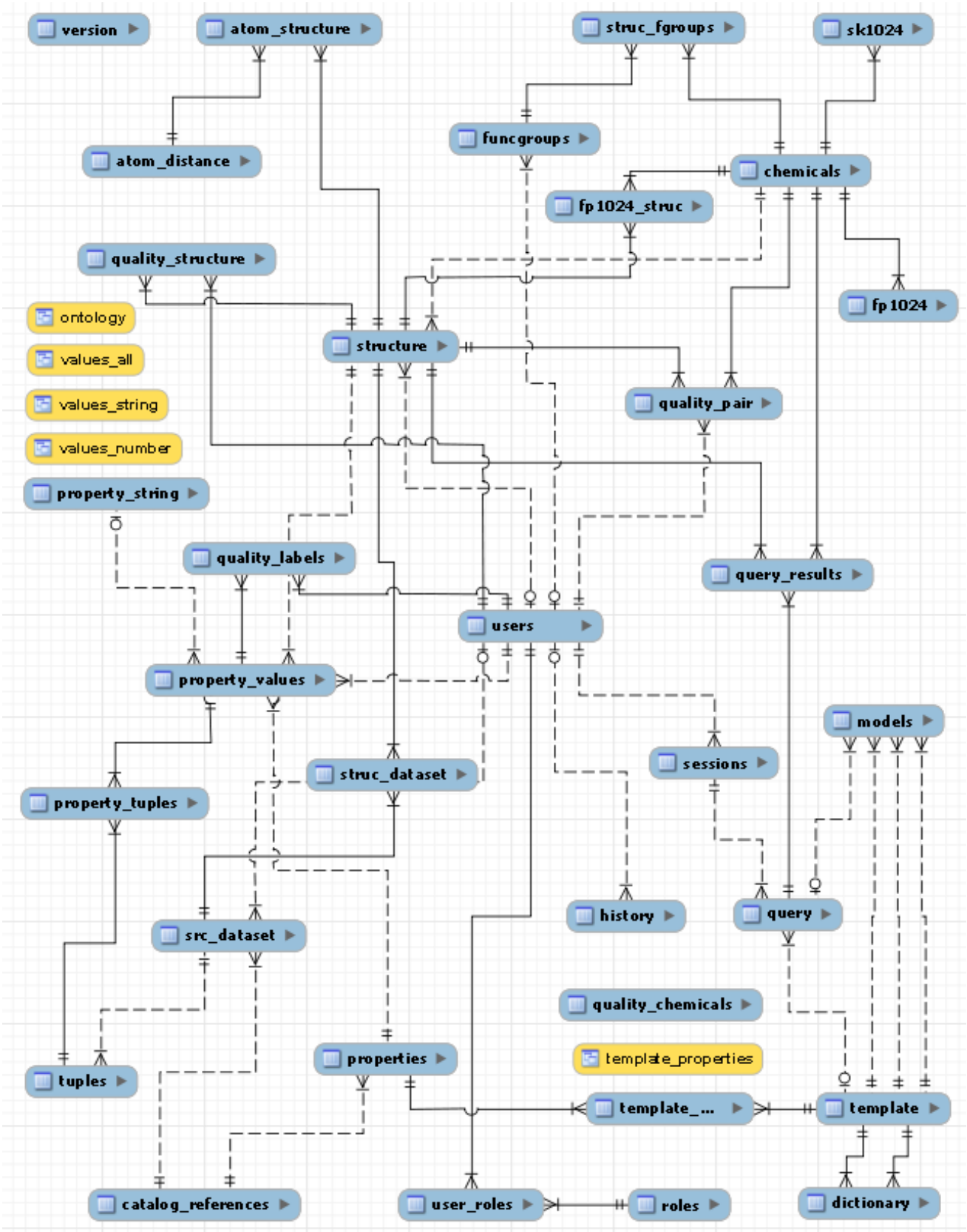

Figure 1 Prototype database entity-relationship diagram

1.2 Chemical Compounds

The chemical compounds are stored in the table *chemicals* and assigned a unique number. If connectivity is available, a unique SMILES[17], as well as InChI[18] and molecular formula is 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 into table *structure* as a compressed text, where supported formats are SDF[19], MOL and CML[20]. 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 not one of the above formats, it is converted to MOL.

1.3 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.

Updates of the chemical structures are recorded and subsequent versions are stored in the *history* table. While importing structures 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) or another structure import step (e.g. updated version of the original file), the new version will be stored and become currently available, while the previous version will be moved to the *history* table.

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.
- Automatically verified, by comparing the structures available under the same chemical compound entry (e.g. imported from different sources) – table *quality_chemicals*:

[17] www.daylight.com/smiles accessed on Feb 14, 2010

[18] www.iupac.org/inchi accessed on Feb 14, 2010

[19] www.mdli.com/downloads/public/ctfile/ctfile.jsp accessed on Feb 14, 2010

[20] en.wikipedia.org/wiki/Chemical_Markup_Language accessed on Feb 14, 2010

- '*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 is a small number of structures, which 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 a comparison.

Furthermore, the results of these 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 in Table 3.

Table 3 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)

In addition, the database provides means for storing different 3D conformations. Recognition of tautomers is under development.

1.4 Identifiers, Descriptors and Properties

The database schema is designed to provide unified storage for arbitrary number of text (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 reference, thus allowing properties with coinciding names, but originating from different sources to be distinguished (e.g. LogP calculated internally by different methods and LogP imported from an external file). Each new property or descriptor is added to a *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 calculation. 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, in order to facilitate queries (e.g. species field, same across all endpoints, in order to be able to search for species). The plan for the future development includes using the alias field to allow mapping to an ontology entry.

Feature/value pairs, which belong to the same experiment, are organized in feature tuples.

The flat list of features provides a flexible storage; however, presenting a long list of properties and descriptors in the user interface might be confusing. Templates (tables *template* and *template_def*) allow to organize properties in groups, as shown in Table 4.

Table 4 Properties grouping templates

Template	Relationship	Parent Template
Endpoints	Top Level Templates	
Identifiers		
Dataset		
Descriptors		
Other	is_a	Endpoint
Ecotoxic effects	is_a	Endpoint
Toxicokinetics	is_a	Endpoint
Environmental fate parameters	is_a	Endpoint
Human health effects	is_a	Endpoint
Physicochemical effects	is_a	Endpoint
Short-term toxicity to algae (inhibition of the exponential growth rate)	is_a	Ecotoxic effects
Toxicity to birds	is_a	Ecotoxic effects
Direct photolysis	is_a	Environmental fate parameters
Oxidation	is_a	Environmental fate parameters
BAF fish	is_a	Bioaccumulation
BAF other organisms	is_a	Bioaccumulation
BCF fish	is_a	Bioconcentration
BCF other organisms	is_a	Bioconcentration
CAS number	is_a	Identifier
RSCBook_Skinsens_dataset.sdf	is_a	Dataset
org.openscience.cdk.qsar.descriptors.molecular.HBondAcceptorCountDescriptor	is_a	Descriptor
org.openscience.cdk.qsar.descriptors.molecular.HBondDonorCountDescriptor	is_a	Descriptor
Verhaar scheme	is_a	Descriptor

Templates themselves can be organized hierarchically, with the help of table *dictionary*. The database is distributed with a set of default templates, including top level templates for Endpoints, Identifiers, Datasets

and Descriptors and a number of endpoints, according to ECHA endpoints classification[21]. Convenience view *ontology*, combines the templates with its hierarchical organisation.

A full SQL dump of the OpenTox prototype database is provided in Appendix B: OpenTox Prototype Database Structure.

2. OpenTox Prototype Database Datasets

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[22]. 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[23] – 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, quote "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 relatively high error rate (estimated at around 10%), because of its paramount importance and relevance in a REACH[24] context;
- Chemical Identifier Resolver[25] – A REST web service evolved from the Chemical Structure Lookup Service[26] and provided by the NCI/CADD group[27] of the US National Cancer Institute[28]. 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 total number of chemical structures known by the service currently is 92,939,226;
- ChemIDplus[29] – 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)[30] databases, including the TOXNET® system[31]. ChemIDplus also provides structure

[21] guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?%20vers=20_08_08 accessed on Feb 14, 2010

[22] www.opentox.org/dev/apis/api-1.1 accessed on Feb 14, 2010

[23] apps.echa.europa.eu/preregistered/pre-registered-sub.aspx accessed on Feb 14, 2010

[24] echa.europa.eu/reach_en.asp accessed on Feb 14, 2010

[25] cactus.nci.nih.gov/chemical/structure accessed on Feb 14, 2010

[26] cactus.nci.nih.gov/cgi-bin/lookup/search accessed on Feb 14, 2010

[27] cactus.nci.nih.gov accessed on Feb 14, 2010

[28] www.cancer.gov accessed on Feb 14, 2010

[29] chem.sis.nlm.nih.gov/chemidplus accessed on Feb 14, 2010

[30] www.nlm.nih.gov accessed on Feb 14, 2010

[31] toxnet.nlm.nih.gov accessed on Feb 14, 2010

searching and direct links to many biomedical resources at NLM and on the Internet for chemicals of interest. The database contains over 388,661 chemical records, of which 295,119 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;

- ChemDraw[32] – a molecule editor developed by the cheminformatics company CambridgeSoft[33]. 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[34] – 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[35], through the ACD/Name to structure Batch[36] software, which converts systematic and non-systematic chemical names of general organic and select biochemical and inorganic compounds into structures;
- ISSCAN[37] – Istituto Superiore di Sanità[38], "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 which 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;
- DSSTox[39] – a project of the US EPA's National Centre for Computational Toxicology[40], 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 OpenTox prototype database the following DSSTox datasets:
    - CPDBAS[41] – 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

[32] www.cambridgesoft.com/software/ChemDraw accessed on Feb 14, 2010

[33] www.cambridgesoft.com accessed on Feb 14, 2010

[34] ecb.jrc.ec.europa.eu/home.php?CONTENU=/DOCUMENTS/QSAR/INFORMATION_SOURCES/EC_CHEMICAL_INVENTORIES accessed on Feb 14, 2010

[35] ecb.jrc.ec.europa.eu/qsar/about-the-group accessed on Feb 14, 2010

[36] www.acdlabs.com/products/name_lab/rename/batch.html accessed on Feb 14, 2010

[37] www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7 accessed on Feb 14, 2010

[38] www.iss.it/chis/?lang=2 accessed on Feb 14, 2010

[39] www.epa.gov/ncct/dsstox accessed on Feb 14, 2010

[40] www.epa.gov/ncct accessed on Feb 14, 2010

[41] www.epa.gov/ncct/dsstox/sdf_cpdbas.html accessed on Feb 14, 2010

6540 chronic, long-term carcinogenesis bioassays reported in 1513 papers in the general literature and 452 Technical Reports of the US National Cancer Institute/National Toxicology Program[42];

- DBPCAN[43] – EPA Water Disinfection By-Products with Carcinogenicity Estimates. The DBPCAN data file, derived from data published by Woo et. al, 2002[44], contains predicted estimates of carcinogenic potential for 209 chemicals detected in finished drinking water samples having undergone water disinfection treatment;
- EPAFHM[45] – EPA Fathead Minnow Acute Toxicity. The EPA Fathead Minnow Acute Toxicity database was generated by the U.S. EPA Mid-Continental Ecology Division (MED)[46] 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)[47] 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 class as most commonly employed in QSAR study. MOA classifications should strengthen the scientific basis for construction of individual QSARs;
- KIERBL[48] – 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)[49]. 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[50], (denoted Group 1), as well as previously unpublished results for an additional 228

[42] ntp.niehs.nih.gov accessed on Feb 14, 2010

[43] www.epa.gov/ncct/dsstox/sdf_dbpcan.html accessed on Feb 14, 2010

[44] 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 Feb 14, 2010

[45] www.epa.gov/ncct/dsstox/sdf_epafhm.html accessed on Feb 14, 2010

[46] www.epa.gov/med accessed on Feb 14, 2010

[47] www.epa.gov/ncct/dsstox/CentralFieldDef.html#MOA accessed on Feb 14, 2010

[48] www.epa.gov/ncct/dsstox/sdf_kierbl.html accessed on Feb 14, 2010

[49] www.epa.gov/scipoly/oscpendo accessed on Feb 14, 2010

[50] 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 Feb 14, 2010

structurally diverse TSCA chemicals[51] for which no ER binding was observed (denoted Group 2);

- IRISTR[52] – 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 noncarcinogenic health effects;
  - Hazard identification, oral slope factors, and oral and inhalation unit risks for carcinogenic effects;
- FDAMDD[53] – FDA Maximum (Recommended) Daily Dose. The US Food and Drug Administration (FDA) Center for Drug Evaluation and Research[54], 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)[55]. 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;

- ECETOC skin irritation[56] – 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)[57] – a database that comprises local lymph node assay (LLNA) data on 211 individual chemicals, encompassing both the chemical and biologic 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 of chemical required to induce a threefold stimulation index. In addition, there are also 42 chemicals that are considered to be nonsensitizers. In terms of chemical diversity, 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 biologic diversity that is known to exist for chemical allergens and non-allergens;

---

[51] www.epa.gov/lawsregs/laws/tsca.html accessed on Feb 14, 2010

[52] www.epa.gov/ncct/dsstox/sdf_iristr.html accessed on Feb 14, 2010

[53] www.epa.gov/ncct/dsstox/sdf_fdamdd.html accessed on Feb 14, 2010

[54] www.fda.gov/Drugs/default.htm accessed on Feb 14, 2010

[55] en.wikipedia.org/wiki/Martindale:_The_Extra_Pharmacopoeia accessed on Feb 14, 2010

[56] ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995).

[57] 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.

- Bioconcentration factor (BCF) Gold Standard Database[58] – a database holding peer reviewed high quality BCF data that is a valuable resource for development of alternative tests.

2.2 Included datasets details

In the following sub-sections we provide further details for each of the 15 OpenTox prototype database datasets.

2.2.1 ECHA list of pre-registered substances

- Name: pre_registered_substances_20090327.xml
- Number of compounds: 143835
- Number of empty structures: 143835
- CAS numbers OK: 118284
- CAS numbers errors: 2
- EINECS numbers OK: 143654
- EINECS numbers errors: 181

2.2.2 Chemical Identifier Resolver

- Name: prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf
- Number of compounds: 85083
- Number of empty structures: 0
- CAS numbers OK: 85083
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.3 ChemIDplus

- Name: prs_complete_20090327-ChemIDplus-20100202
- Number of compounds: 78525
- Number of empty structures: 1
- CAS numbers OK: 78525
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.4 ChemDraw

- Name: ECHA-ChemDraw

[58] ambit.sourceforge.net/euras accessed on Feb 14, 2010

- Number of compounds: 20052
- Number of empty structures: 0
- CAS numbers OK: 0
- CAS numbers errors: 0
- EINECS numbers OK: 20051
- EINECS numbers errors: 0

2.2.5 JRC PRS list

- Name: PRS_processed_file-VJ.sdf
- Number of compounds: 80410
- Number of empty structures: 0
- CAS numbers OK: 80410
- CAS numbers errors: 0
- EINECS numbers OK: 70749
- EINECS numbers errors: 0

2.2.6 ISSCAN

- Name: ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA
- Number of compounds: 1153
- Number of empty structures: 12
- CAS numbers OK: 1140
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.7 CPDBAS

- Name: CPDBAS: Carcinogenic Potency Database Summary Tables - All Species
- Number of compounds: 1547
- Number of empty structures: 39
- CAS numbers OK: 1527
- CAS numbers errors: 0
- EINECS numbers OK: 0
- EINECS numbers errors: 0

2.2.8 DBPCAN

- Name: DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates
- 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

2.2.9 EPAFHM

- Name: EPAFHM: EPA Fathead Minnow Acute Toxicity
- 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

2.2.10 KIERBL

- Name: KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.)
- 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

2.2.11 IRISTR

- Name: IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data
- 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

2.2.12 FDAMDD

- Name: FDAMDD: FDA Maximum (Recommended) Daily Dose
- 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 ECETOC skin irritation

- Name: ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)
- 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

2.2.14 Skin sensitisation (LLNA)

- Name: Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives
- 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

2.2.15 Bioconcentration factor (BCF) Gold Standard Database

- Name: Bioconcentration factor (BCF) Gold Standard Database
- Number of compounds: 1130
- Number of empty structures: 1130
- CAS numbers OK: 1121
- CAS numbers errors: 7
- 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 5. The first column indicates whether we have 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 5 OpenTox prototype database automatic classification results distribution

Label	Details	Number of Chemicals
Consensus	14	1
Consensus	11	1
Consensus	10	1
Consensus	9	8
Consensus	8	36
Consensus	7	69
Consensus	6	131
Consensus	5	675
Consensus	4	16870
Consensus	3	40989
Consensus	2	9687
Majority	1:7	1
Majority	1:6	2
Majority	1:5	11
Majority	1:4	56
Majority	1:3	1287
Majority	1:2	7389
Majority	2:5	3
Majority	2:4	2
Majority	2:3	7
Majority	1:1:4	1
Majority	1:1:3	16
Majority	1:1:2	191
Majority	1:2:2	1
Majority	1:1:1:2	5
Unconfirmed	1	14975

Label	Details	Number of Chemicals
Ambiguous	1:1	2624
Ambiguous	2:2	427
Ambiguous	1:1:1	519
Ambiguous	1:1:2	1
Ambiguous	1:1:1:1	15

In order 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 6 lists the results and highlights the fact that ChemIDplus (1.24% error rate) seems to provide better quality data when compared to other generic data sources such as the Chemical Identifier Resolver data (4.75% error rate) and JRC PRS list data (5.83% error rate). All DSSTox data sources prove to have excellent quality (0% or close to 0% error rate).

Table 6 OpenTox prototype database quality labels distribution

Dataset	OK	Probably OK	Probably ERROR	Unknown	Probably ERROR%
ECHA list of pre-registered substances	N/A	N/A	N/A	N/A	N/A
Chemical Identifier Resolver	67779	5314	3638	3471	4.75%
ChemIDplus	64802	7986	921	1745	1.24%
ChemDraw	17918	1147	502	478	2.57%
JRC PRS list	61332	4833	4022	2880	5.83%
ISSCAN	931	50	98	62	9.40%
CPDBAS	778	37	0	693	0%
DBPCAN	60	2	0	147	0%
EPAFHM	281	5	0	331	0%
KIERBL	102	1	0	175	0%
IRISTR	346	16	0	177	0%
FDAMDD	213	19	1	983	0.08%
ECETOC skin irritation	158	12	0	5	0%
Skin sensitisation (LLNA)	160	7	4	38	1.95%
Bioconcentration factor (BCF) Gold Standard Database	N/A	N/A	N/A	N/A	N/A

ISSCAN's relatively high error rate (9.40%) is probably mainly due to removal of disconnected parts of structures, which is not duly reported (e.g. salts removal should be acknowledged in the database at

ontological level). For instance, a relevant example can been seen at
http://apps.ideaconsult.net:8180/ambit2/compound/16090/conformer

As you can see in Figure 2 we have a majority of 4 identical structures, compared to one outlier, which happens to come from ISSCAN and is stripped of the 3 Na^+ ions. However, the corresponding acid has a different chemical name and registry identifiers (CAS, EINECS), as highlighted by the following comparison:

ISSCAN: http://apps.ideaconsult.net:8180/ambit2/compound/16090/conformer/410569

CASRN 1934-21-0

ChemicalName trisodium 5-hydroxy-1-(4-sulphophenyl)-4-(4-sulphophenylazo)pyrazole-3-carboxylate

ChemicalName ACID YELLOW-23

ChemicalName C16H9N4Na3O9S2

ChemicalName FD & C yellow no. 5

EINECS 217-699-5

Versus: http://apps.ideaconsult.net:8180/ambit2/compound/47131

CASRN 34175-08-1

ChemicalName 4,5-dihydro-5-oxo-1-(4-sulphophenyl)-4-[(4-sulphophenyl)azo]-1H-pyrazole-3-carboxylic acid

ChemicalName C16H12N4O9S2

EINECS 251-862-1

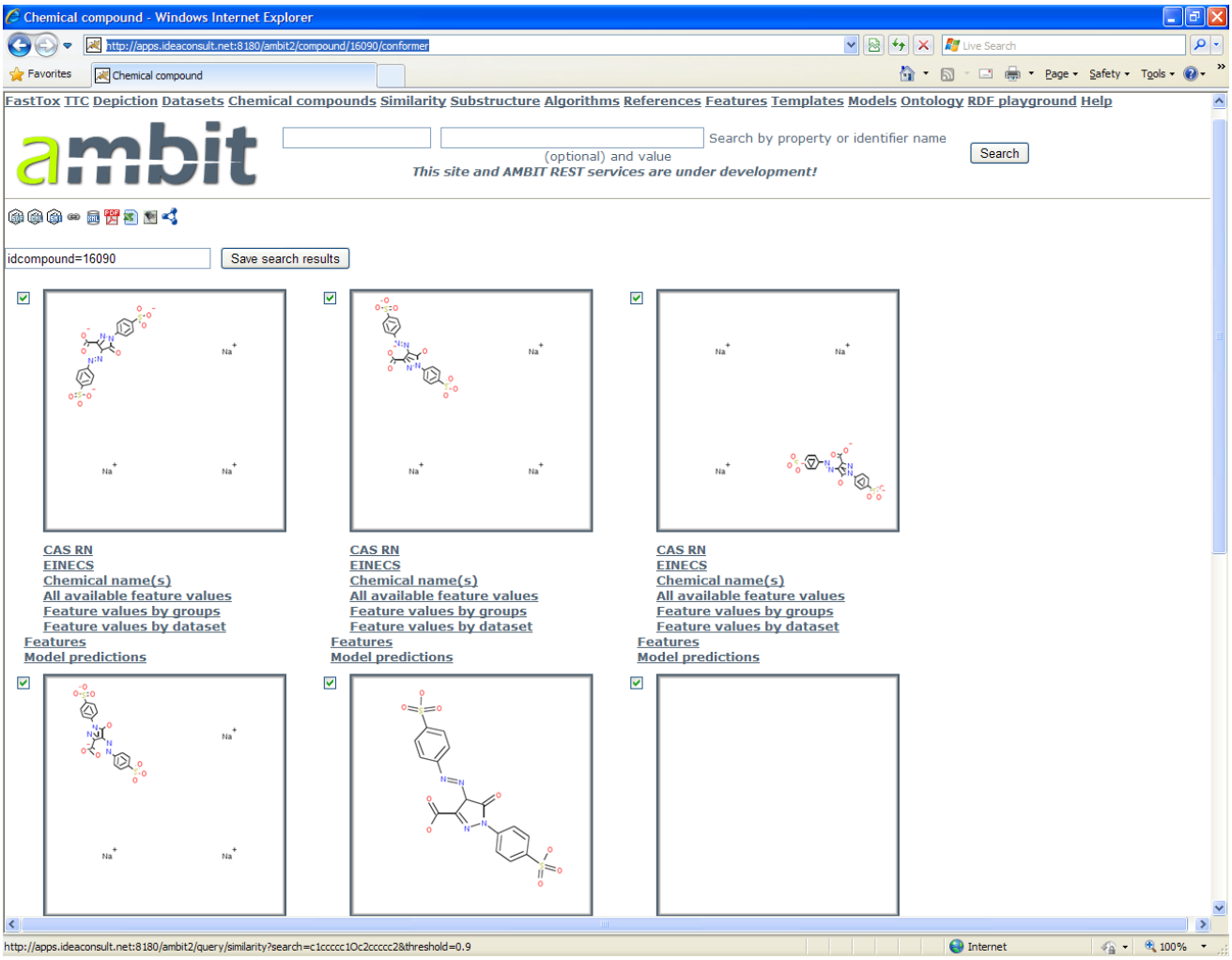

Figure 2 ISSCAN unacknowledged salt removal

It is well known that in many cases salts removal is a required pre-processing step for QSAR modelling. However, in order to avoid similar problems either this pre-processing should be done on-the-fly or the database (and its associated ontology) should indicate that such a step has been applied and provide means to obtain the original structure. In ISSCAN's case neither of these options is available, which results in the reported relatively high error rate. This also highlights the need for more detailed ontologies development.

The above mentioned 15 datasets included in the OpenTox prototype database are accessible online at http://apps.ideaconsult.net:8180/ambit2/dataset through the OpenTox REST API[59]. In particular, these datasets are available through the OpenTox Dataset API[60] and could be accessed through 3rd party web services and/or client implementing this API. They could be retrieved as complete datasets or via various kinds of search operations over chemical names, registry identifiers, structures, sub-structures (SMARTS)[61], similarity, Daylight SMILES[62], InChI[63], etc. The implemented API also allows adding new datasets through POST operations.

[59] www.opentox.org/dev/apis/api-1.1 accessed on Feb 14, 2010

[60] www.opentox.org/dev/apis/api-1.1/dataset accessed on Feb 14, 2010

[61] www.daylight.com/dayhtml/doc/theory/theory.smarts.html accessed on Feb 4, 2010

[62] www.daylight.com/smiles accessed on Feb 14, 2010

All these functionalities are used by the OpenTox ToxPredict web application, available at toxpredict.org. The full raw database statistics as reported by AmbitXT are provided in Appendix A: OpenTox Prototype Database Raw Statistics.

3. OpenTox Toxicological Endpoints Ontology

We have developed an OWL (Web Ontology Language)[64] ontology of toxicological endpoints[65], which corresponds to the endpoint classification of REACH guidance documents[66] and allows for unique mapping between endpoints from various inventories. Figure 3 provides a glimpse at the ECHA endpoints ontology summary graph, while Figure 4 illustrates its human health specific part.

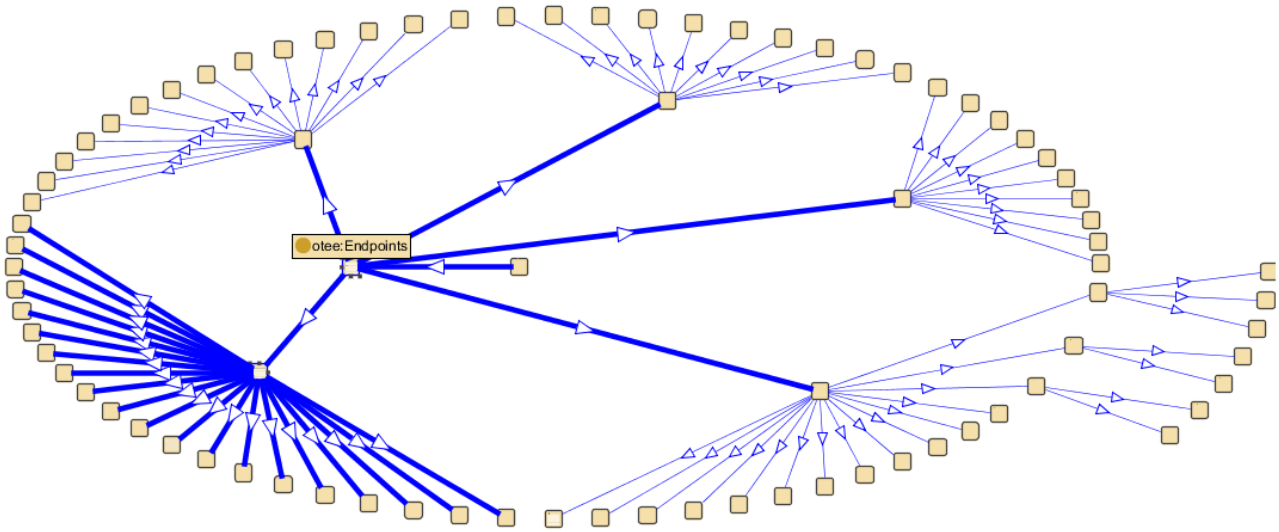

Figure 3 ECHA endpoints ontology summary graph

(Q)SAR model quality crucially depends on the clarity of endpoints and experimental protocol 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 which is in line with current ECHA REACH guidance. Using this ontology, each attribute in a toxicological dataset can be associated with an entry to the ontology, therefore allowing a unique mapping between endpoints in various and heterogeneous datasets. The mapping of chemical compound properties,

---

[63] www.iupac.org/inchi accessed on Feb 14, 2010

[64] www.w3.org/TR/owl-features accessed on Feb 14, 2010

[65] opentox.org/data/documents/development/RDF%20files/Endpoints/view accessed on Feb 27, 2010

[66] guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?%20vers=20_08_08 accessed on Feb 14, 2010

stored in the OpenTox prototype 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.

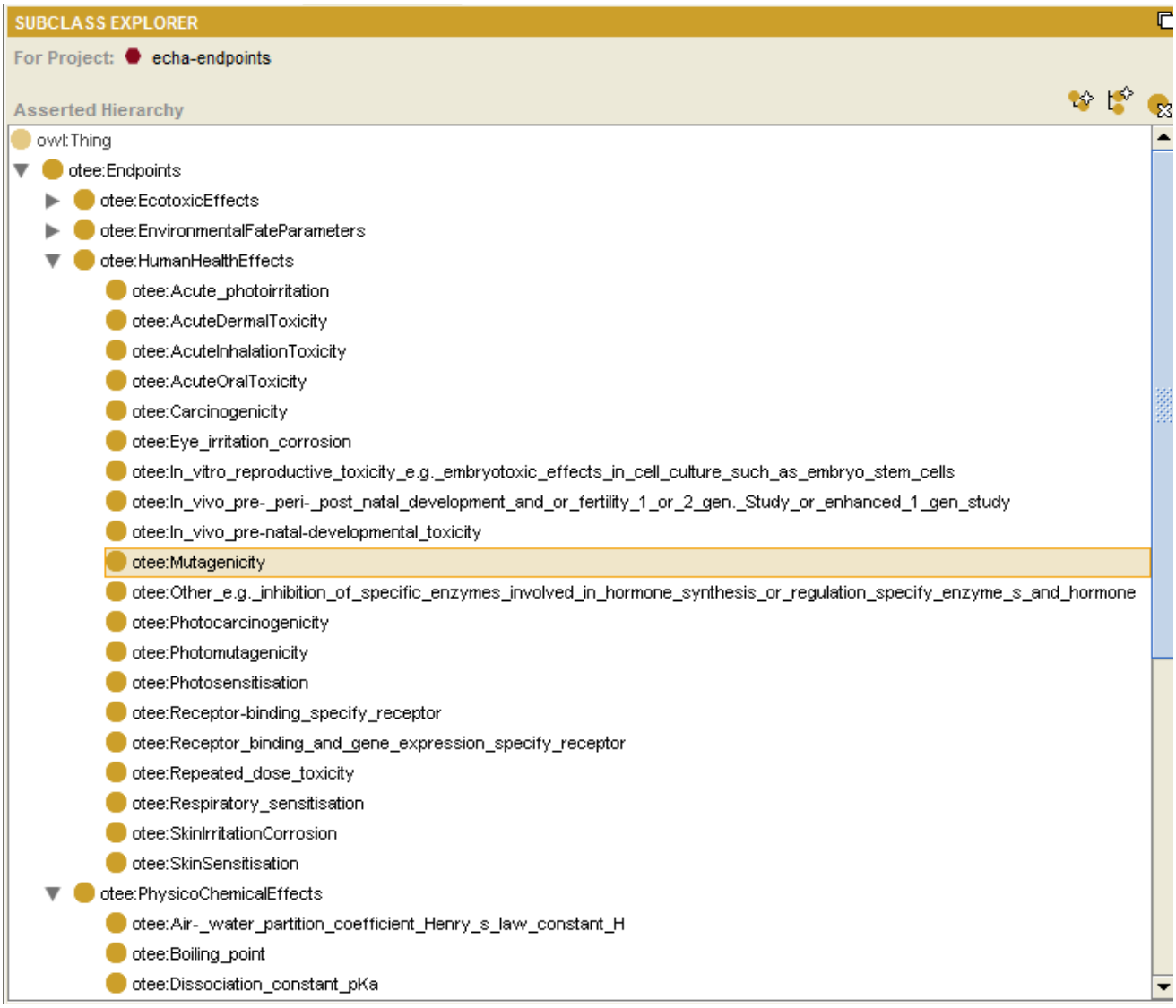

Figure 4 Human health specific part of the ECHA endpoints ontology

A full OWL dump of the Toxicological Endpoints Ontology is provided in Appendix C: OpenTox Toxicological Endpoints Ontology. The ontology has been used for mapping the relevant OpenTox prototype 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[67] stores the endpoints ontology, along with other OpenTox-relevant ontologies (opentox.owl[68], algorithm types[69] and Blue Obelisk Descriptor Ontology[70]),

[67] http://apps.ideaconsult.net:8180/ontology accessed on Feb 25, 2010

[68] opentox.org/data/documents/development/RDF%20files/OpenToxOntology/view accessed on Feb 25, 2010

as well as dynamic information of available models, algorithms and features, provided by OpenTox services. The ToxPredict application[71] (step 3) queries[72] the Ontology service by standard SPARQL[73] interface for all available models, and retrieves associated information about endpoint modelled, algorihtms used to create models and independent and target variables used in models.

4. Further Work Directions

During the second half of the OpenTox project we're planning to:

- identify and include more datasets in the OpenTox database, satisfying the established rigorous quality criteria, e.g.:
    - Fraunhofer ITEM RepDose[74]: subacute to chronic toxicity data for 655 chemicals, retrieved from around 2280 studies, carried out in rats, mice or dogs with oral or inhalation exposure. Each chemical is characterised by at least one, but up to 15 studies, with the majority of chemicals having 1 to 4 studies. Provides a sound basis for the analysis of relationships between chemical functional groups/categories and target organs in repeated dose studies;
    - ISS ISSMIC – 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 three (well ahead of rodent caricnogenicity) as 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 and chemical structures, and is the basis for establishing sound read-across and QSAR risk assessment;
    - IBMC data on sensitization, reproductive toxicity, embryotoxicity and teratogenicity for several thousands of chemicals, along with relevant physico-chemical and structural information for these chemicals.
- perform further work on data curation in order to resolve as many inconsistencies and errors as possible. An obvious target for this work would be compounds which have one of the following automatic classification quality labels assigned in the OpenTox Prototype Database (in decreasing order of importance):
    1. Ambiguous;
    2. Unconfirmed;
    3. Majority.

---

[69] opentox.org/data/documents/development/RDF%20files/AlgorithmTypes/view accessed on Feb 25, 2010

[70] opentox.org/data/documents/development/RDF%20files/BlueObeliskOntology/view accessed on Feb 25, 2010

[71] toxpredict.org accessed on Feb 25, 2010

[72] opentox.org/data/documents/development/RDF%20files/JavaOnly/query-reasoning-with-jena-and-sparql accessed on Feb 25, 2010

[73] http://www.w3.org/TR/rdf-sparql-query/ accessed on Feb 25, 2010

[74] www.fraunhofer-repdose.de accessed on February 23, 2010

- implement OpenTox API-compliant workflows for batch online data retrieval from 3rd party sources (e.g. Chemical Identifier Resolver, ChemIDplus, PubChem, IUCLID5[75], ToxCast[76], etc);
- apply the ontologies, under development by partner ISS, for more detailed representation of links between various dataset fields, endpoints and models;
- introduce authentication and authorization in the OpenTox framework, allowing for confidential and restricted access to certain datasets, as well as per-user workspace and profile management.

5. Conclusions

The OpenTox prototype 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 webservices. 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 for online beta-testing at toxpredict.org, illustrates one generic use case for the OpenTox prototype database. Users can either search the OpenTox prototype database, which includes currently quality-labelled data for ~150,000 chemicals, grouped in more than a dozen datasets, or can upload their own chemical structure data and eventually chose to allow third parties to work on it in a collaborative way. At the time of this writing, ToxPredict provides access to 9 ready-to-use models, addressing 8 different endpoints. It allows transparent and sound scientific modelling of chemical properties, in compliance with the best current practices for data management and records keeping, implemented in the OpenTox prototype database.

Another important aspect of the OpenTox prototype 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, in Intranets (or even offline), when very tight security policies would have to be met. Third parties, willing to deploy the OpenTox prototype database in-house, could select a relevant subset of datasets to install, 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 the 21st century.

[75] iuclid.echa.europa.eu accessed on Feb 14, 2010

[76] www.epa.gov/ncct/toxcast accessed on Feb 14, 2010

Appendix A: OpenTox Prototype Database Raw Statistics

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

Number of datasets 15

Dataset	Number of compounds
pre_registered_substances_20090327.xml	143835
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	85083
prs_complete_20090327-ChemIDplus-20100202	78525
ECHA-ChemDraw	20052
PRS_processed_file-VJ.sdf	80410
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	176
Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives	209
Bioconcentration factor (BCF) Gold Standard Database	1130
ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA	1153
CPDBAS: Carcinogenic Potency Database Summary Tables - All Species	1547
DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates	209
EPAFHM: EPA Fathead Minnow Acute Toxicity	617
KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.)	278
IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data	544
FDAMDD: FDA Maximum (Recommended) Daily Dose	1216

Dataset	Number of empty structures
pre_registered_substances_20090327.xml	143835
prs_complete_20090327-ChemIDplus-20100202	1
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	1
Bioconcentration factor (BCF) Gold Standard Database	1130
ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA	12
CPDBAS: Carcinogenic Potency Database Summary Tables - All Species	39
IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data	5

Quality labels summary Entire database

Label	Details	Number of chemicals
Consensus	10	1
Consensus	11	1
Consensus	14	1

Consensus	2	9687
Consensus	3	40989
Consensus	4	16870
Consensus	5	675
Consensus	6	131
Consensus	7	69
Consensus	8	36
Consensus	9	8
Majority	1:1:1:2	5
Majority	1:1:2	191
Majority	1:1:3	16
Majority	1:1:4	1
Majority	1:2	7389
Majority	1:2:2	1
Majority	1:3	1287
Majority	1:4	56
Majority	1:5	11
Majority	1:6	2
Majority	1:7	1
Majority	2:3	7
Majority	2:4	2
Majority	2:5	3
Unconfirmed	1	14975
Ambiguous	1:1	2624
Ambiguous	1:1:1	519
Ambiguous	1:1:1:1	15
Ambiguous	1:2:2	1
Ambiguous	2:2	427

Dataset	Mode	label	Number of compounds
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	comparison	OK	67779
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	comparison	ProbablyOK	5314
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	comparison	Unknown	3471
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	comparison	ProbablyERROR	3638
prs_complete_20090327-ChemIDplus-20100202	comparison	OK	64802
prs_complete_20090327-ChemIDplus-20100202	comparison	ProbablyOK	7986
prs_complete_20090327-ChemIDplus-20100202	comparison	Unknown	1745
prs_complete_20090327-ChemIDplus-20100202	comparison	ProbablyERROR	921
ECHA-ChemDraw	comparison	OK	17918
ECHA-ChemDraw	comparison	ProbablyOK	1147
ECHA-ChemDraw	comparison	Unknown	478
ECHA-ChemDraw	comparison	ProbablyERROR	502
PRS_processed_file-VJ.sdf	comparison	OK	61332
PRS_processed_file-VJ.sdf	comparison	ProbablyOK	4833
PRS_processed_file-VJ.sdf	comparison	Unknown	2880
PRS_processed_file-VJ.sdf	comparison	ProbablyERROR	4022

ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison OK 158

ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison ProbablyOK 12

ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995) comparison Unknown 5

Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives comparison OK 160

Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives comparison ProbablyOK 7

Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives comparison Unknown 38

Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives comparison ProbablyERROR 4

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA comparison OK 931

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA comparison ProbablyOK 50

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA comparison Unknown 62

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA comparison ProbablyERROR 98

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species comparison OK 778

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species comparison ProbablyOK 37

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species comparison Unknown 693

DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates comparison OK 60

DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates comparison ProbablyOK 2

DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates comparison Unknown 147

EPAFHM: EPA Fathead Minnow Acute Toxicity comparison OK 281

EPAFHM: EPA Fathead Minnow Acute Toxicity comparison ProbablyOK 5

EPAFHM: EPA Fathead Minnow Acute Toxicity comparison Unknown 331

KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.) comparison OK 102

KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.) comparison ProbablyOK 1

KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.) comparison Unknown 175

IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data comparison OK 346

IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data comparison ProbablyOK 16

IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data comparison Unknown 177

FDAMDD: FDA Maximum (Recommended) Daily Dose comparison OK 213

FDAMDD: FDA Maximum (Recommended) Daily Dose comparison ProbablyOK 19

FDAMDD: FDA Maximum (Recommended) Daily Dose comparison Unknown 983

FDAMDD: FDA Maximum (Recommended) Daily Dose comparison ProbablyERROR 1

Datasets	Mode	label	Number of compounds
pre_registered_substances_20090327.xml	CAS numbers	OK	118284
pre_registered_substances_20090327.xml	CAS numbers	ERROR	2

pre_registered_substances_20090327.xml EINECS numbers OK 143654

pre_registered_substances_20090327.xml EINECS numbers ERROR 181

prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf CAS numbers OK 85083

prs_complete_20090327-ChemIDplus-20100202 CAS numbers OK 78525

ECHA-ChemDraw EINECS numbers OK 20051

PRS_processed_file-VJ.sdf CAS numbers OK 80410

PRS_processed_file-VJ.sdf EINECS numbers OK 70749

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

Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives CAS numbers OK 203

Bioconcentration factor (BCF) Gold Standard Database CAS numbers OK 1121

Bioconcentration factor (BCF) Gold Standard Database CAS numbers ERROR 7

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA CAS numbers OK 1140

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species CAS numbers OK 1527

DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates CAS numbers OK 179

EPAFHM: EPA Fathead Minnow Acute Toxicity CAS numbers OK 617

KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.) CAS numbers OK 278

IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data CAS numbers OK 536

FDAMDD: FDA Maximum (Recommended) Daily Dose CAS numbers OK 1216

Dataset Mode label Number of compounds

pre_registered_substances_20090327.xml CAS numbers,EINECS numbers OK 143654

pre_registered_substances_20090327.xml EINECS numbers,CAS numbers ERROR 183

prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf CAS numbers OK 85083

prs_complete_20090327-ChemIDplus-20100202 CAS numbers OK 78525

ECHA-ChemDraw EINECS numbers OK 20051

PRS_processed_file-VJ.sdf CAS numbers,EINECS numbers OK 80410

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

Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives CAS numbers OK 203

Bioconcentration factor (BCF) Gold Standard Database CAS numbers OK 1121

Bioconcentration factor (BCF) Gold Standard Database CAS numbers ERROR 7

ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA CAS numbers OK 1140

CPDBAS: Carcinogenic Potency Database Summary Tables - All Species CAS numbers OK 1527

DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates CAS numbers OK 179

EPAFHM: EPA Fathead Minnow Acute Toxicity CAS numbers OK 617

KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.) CAS numbers OK 278

IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data CAS numbers OK 536

FDAMDD: FDA Maximum (Recommended) Daily Dose CAS numbers OK 1216

Datasets	Mode	label	Number of compounds
pre_registered_substances_20090327.xml,prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf,prs_complete_20090327-ChemIDplus-20100202,ECHA-ChemDraw,PRS_processed_file-VJ.sdf,ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995),Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives,Bioconcentration factor (BCF) Gold Standard Database,ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA,CPDBAS: Carcinogenic Potency Database Summary Tables - All Species,DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates,EPAFHM: EPA Fathead Minnow Acute Toxicity,KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.),IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data,FDAMDD: FDA Maximum (Recommended) Daily Dose	CAS numbers,EINECS numbers	OK	146905
pre_registered_substances_20090327.xml,ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995),Bioconcentration factor (BCF) Gold Standard Database	EINECS numbers,CAS numbers	ERROR	190

Dataset	Structure type	Number of compounds
pre_registered_substances_20090327.xml	NA	143835
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	3D no H	3647
prs_complete_20090327-CSLS-StructByCAS-MERGED.sdf	3D with H	81436
prs_complete_20090327-ChemIDplus-20100202	NA	1
prs_complete_20090327-ChemIDplus-20100202	MARKUSH	2848
prs_complete_20090327-ChemIDplus-20100202	3D no H	71205
prs_complete_20090327-ChemIDplus-20100202	3D with H	4471
ECHA-ChemDraw	3D no H	19036
ECHA-ChemDraw	3D with H	1016
PRS_processed_file-VJ.sdf	MARKUSH	87
PRS_processed_file-VJ.sdf	3D no H	77508
PRS_processed_file-VJ.sdf	3D with H	2815
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)	3D no H	171
ECETOC Technical Report No. 66 Skin irritation and corrosion Reference Chemicals data base (1995)	3D with H	4
Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives	3D no H	2
Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives	3D with H	207
Bioconcentration factor (BCF) Gold Standard Database	NA	1130
ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA	NA	12
ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA	3D no H	1127
ISSCAN: Istituto Superiore di Sanita, CHEMICAL CARCINOGENS: STRUCTURES AND EXPERIMENTAL DATA	3D with H	14
CPDBAS: Carcinogenic Potency Database Summary Tables - All Species	NA	39
CPDBAS: Carcinogenic Potency Database Summary Tables - All Species	3D no H	1422
CPDBAS: Carcinogenic Potency Database Summary Tables - All Species	3D with H	86
DBPCAN: EPA Water Disinfection By-Products with Carcinogenicity Estimates	3D no H	209

```
EPAFHM: EPA Fathead Minnow Acute Toxicity    3D no H 606
EPAFHM: EPA Fathead Minnow Acute Toxicity    3D with H      11
KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.)        3D no H 276
KIERBL: EPA Estrogen Receptor Ki Binding Study (Laws et al.)        3D with H      2
IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data NA      5
IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data 3D no H 532
IRISTR: EPA Integrated Risk Information System (IRIS) Toxicity Review Data 3D with H      7
FDAMDD: FDA Maximum (Recommended) Daily Dose 3D no H 984
FDAMDD: FDA Maximum (Recommended) Daily Dose 3D with H      232
```

Appendix B: OpenTox Prototype Database Structure

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

```
SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='TRADITIONAL';


-- -----------------------------------------------------
-- Table `roles`  User roles
-- -----------------------------------------------------
DROP TABLE IF EXISTS `roles`;
CREATE TABLE  `roles` (
  `role_name` varchar(16) character set utf8 collate utf8_bin NOT NULL,
  PRIMARY KEY  (`role_name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

-- -----------------------------------------------------
-- Table `users` Users
-- -----------------------------------------------------
DROP TABLE IF EXISTS `users`;
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;

-- -----------------------------------------------------
-- Table `roles` Roles assigned to users
-- -----------------------------------------------------
DROP TABLE IF EXISTS `user_roles`;
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_2` FOREIGN KEY (`role_name`) REFERENCES `roles` (`role_name`) ON DELETE
CASCADE ON UPDATE CASCADE,
  CONSTRAINT `FK_user_roles_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`) ON DELETE
CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8;


-- ----------------------------------------------------
-- Table `references`
-- ----------------------------------------------------
DROP TABLE IF EXISTS `catalog_references`;
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,
  PRIMARY KEY  (`idreference`),
  UNIQUE KEY `Index_2` (`title`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

-- ----------------------------------------------------
-- Table `chemicals`
-- ----------------------------------------------------
DROP TABLE IF EXISTS `chemicals`;
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,
  `hashcode` bigint(20) NOT NULL default '0',
  `label` enum('OK','UNKNOWN','ERROR') NOT NULL default 'UNKNOWN',
  PRIMARY KEY  (`idchemical`),
  KEY `sinchi` (`inchi`(760)),
  KEY `ssmiles` (`smiles`(760)),
  KEY `idchemical` (`idchemical`),
  KEY `inchi` (`inchi`(767)),
  KEY `formula` (`formula`),
  KEY `hashcode` USING BTREE (`hashcode`),
  KEY `Index_8` (`label`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

-- ----------------------------------------------------
-- Table `structure`
-- ----------------------------------------------------
DROP TABLE IF EXISTS `structure` ;
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') collate utf8_bin NOT NULL default 'CML',
  `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,
  PRIMARY KEY  (`idstructure`),
  KEY `FK_structure_2` (`user_name`),
  KEY `idchemical` USING BTREE (`idchemical`),
  KEY `Index_4` (`label`),
  KEY `Index_5` (`idstructure`,`user_name`),
  CONSTRAINT `fk_idchemical` FOREIGN KEY (`idchemical`) REFERENCES `chemicals` (`idchemical`) ON
DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `FK_structure_2` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`) ON DELETE
SET NULL ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


DELIMITER $
CREATE TRIGGER copy_history BEFORE UPDATE ON structure
  FOR EACH ROW BEGIN
   INSERT INTO history (idstructure,structure,format,updated,user_name,type_structure,label)
        SELECT idstructure,structure,format,updated,user_name,type_structure,label FROM structure
        WHERE structure.idstructure = OLD.idstructure;
  END $
DELIMITER ;


-- ----------------------------------------------------
-- Table `properties`
-- ----------------------------------------------------
DROP TABLE IF EXISTS `properties`;
CREATE TABLE  `properties` (
  `idproperty` int(10) unsigned NOT NULL auto_increment,
  `idreference` int(11) unsigned NOT NULL default '0',
  `name` varchar(128) collate utf8_bin NOT NULL default '',
  `units` varchar(16) collate utf8_bin NOT NULL default '',
  `comments` varchar(128) collate utf8_bin NOT NULL default '',
  `islocal` tinyint(1) NOT NULL default '0',
  PRIMARY KEY  USING BTREE (`idproperty`),
  UNIQUE KEY `ddictionary_name` USING BTREE (`name`,`idreference`),
  KEY `ddictionary_idref` (`idreference`),
  CONSTRAINT `FK_properties_1` FOREIGN KEY (`idreference`) REFERENCES `catalog_references`
(`idreference`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


-- ----------------------------------------------------
```

```
-- Table `property_string` string values
-- ----------------------------------------------------
DROP TABLE IF EXISTS `property_string`;
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 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


-- ----------------------------------------------------
-- Table `template` defines templates
-- ----------------------------------------------------
DROP TABLE IF EXISTS `template`;
CREATE TABLE  `template` (
  `idtemplate` int(10) unsigned NOT NULL auto_increment,
  `name` varchar(128) collate utf8_bin default NULL,
  PRIMARY KEY  (`idtemplate`),
  UNIQUE KEY `template_list_index4157` USING BTREE (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


--
-- Dumping data for table `template`
--


LOCK TABLES `template` WRITE;
/*!40000 ALTER TABLE `template` DISABLE KEYS */;
INSERT INTO `template` VALUES (84,NULL),(7,'Acute dermal toxicity'),(8,'Acute inhalation
toicity'),(9,'Acute oral toxicity'),(10,'Acute photoirritation'),(11,'Acute toxicity to fish
(lethality)'),(12,'Adsorption/Desorption in sediment'),(13,'Adsorption/Desorption in soil
'),(14,'Air- water partition coefficient (Henry`s law constant, H)'),(15,'BAF fish'),(16,'BAF other
organisms '),(17,'BCF fish'),(18,'BCF other organisms '),(78,'Bioaccumulation'),(79,'Bioconcentration
'),(19,'Biodegradation time frame (primary, ultimate degradation)'),(20,'Blood-brain barrier
penetration'),(21,'Blood-lung barrier penetration'),(22,'Blood-testis barrier
penetration'),(23,'Boiling point'),(89,'CasRN'),(24,'Carcinogenicity'),(91,'Names'),(27,'DNA-
binding'),(86,'Descriptors'),(25,'Direct photolysis'),(26,'Dissociation constant (pKa)'),(1,'Ecotoxic
effects'),(83,'Endocrine Activity'),(85,'Endpoints'),(2,'Environmental fate parameters '),(28,'Eye
irritation/corrosion'),(29,'Gastrointestinal absorption'),(3,'Human health effects'),(30,'Hydrolysis
'),(90,'IUPAC name'),(88,'Identifiers'),(31,'In vitro reproductive toxicity (e.g. embryotoxic effects
in cell culture such as embryo stem cells)  '),(32,'In vivo pre-, peri-, post natal development and /
or fertility (1 or 2 gen. Study or enhanced 1 gen study) '),(33,'In vivo pre-natal-developmental
toxicity'),(34,'Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)'),(35,'Long-
term toxicity (survival, growth, reproduction)'),(36,'Long-term toxicity to Daphnia (lethality,
inhibition of reproduction)'),(37,'Long-term toxicity to fish (egg/sac fry, growth inhibition of
juvenile fish, early life stage, full life cycle)'),(38,'Melting point'),(39,'Metabolism (including
metabolic clearance)'),(40,'Microbial inhibition (activated sludge respiration inhibition, inhibition
of nitrification, other)'),(41,'Mutagenicity '),(42,'Octanol-air partition coefficient
(Koa)'),(43,'Octanol-water distribution coefficient (D)'),(44,'Octanol-water partition coefficient
(Kow)'),(45,'Ocular membrane penetration'),(46,'Organic carbon-sorption partition coefficient
(organic carbon; Koc)'),(4,'Other'),(47,'Other (e.g. inhibition of specific enzymes involved in
hormone synthesis or regulation, specify enzyme(s) and hormone)'),(48,'Oxidation '),(80,'Persistence:
Abiotic degradation in air (Phototransformation)'),(81,'Persistence: Abiotic degradation in
water'),(82,'Persistence:
Biodegradation'),(49,'Photocarcinogenicity'),(50,'Photomutagenicity'),(51,'Photosensitisation'),(5,'P
hysicochemical effects '),(52,'Placental barrier penetration'),(53,'Protein-binding'),(54,'Ready/not
ready biodegradability'),(55,'Receptor binding and gene expression (specify
```

```
receptor)'),(56,'Receptor-binding (specify receptor)'),(57,'Repeated dose toxicity
'),(58,'Respiratory sensitisation'),(59,'Short term toxicity (feeding, gavage, other)'),(61,'Short-
term toxicity to Daphnia (immobilisation)'),(60,'Short-term toxicity to algae (inhibition of the
exponential growth rate)'),(62,'Skin irritation /corrosion'),(63,'Skin penetration'),(64,'Skin
sensitisation'),(65,'Surface tension'),(77,'Toxicity to birds'),(66,'Toxicity to earthworms
(survival, growth, reproduction)'),(67,'Toxicity to plants (leaves, seed germination, root
elongation)'),(68,'Toxicity to sediment organisms (survival, growth, reproduction)'),(69,'Toxicity to
soil invertebrates (survival, growth, reproduction)'),(70,'Toxicity to soil microorganisms
(inhibition of C-mineralisation, inhibition of N-mineralisation, other)'),(6,'Toxicokinetics
'),(71,'Vapour pressure'),(72,'Vegetation-air partition coefficient'),(73,'Vegetation-soil partition
coefficient'),(74,'Vegetation-water partition coefficient'),(75,'Water solubility'),(92,'Dataset');

/*!40000 ALTER TABLE `template` ENABLE KEYS */;

UNLOCK TABLES;


-- ------------------------------------------------------

-- Table `template_def` template definitions

-- ------------------------------------------------------

DROP TABLE IF EXISTS `template_def`;

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  USING BTREE (`idtemplate`,`idproperty`),

  KEY `FK_template_def_2` (`idproperty`),

  CONSTRAINT `FK_template_def_2` FOREIGN KEY (`idproperty`) REFERENCES `properties` (`idproperty`) ON
DELETE CASCADE ON UPDATE CASCADE,

  CONSTRAINT `FK_template_def_1` FOREIGN KEY (`idtemplate`) REFERENCES `template` (`idtemplate`) ON
DELETE CASCADE ON UPDATE CASCADE

) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;



-- ------------------------------------------------------

-- Table `dictionary`

-- ------------------------------------------------------

DROP TABLE IF EXISTS `dictionary`;

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_2` FOREIGN KEY (`idobject`) REFERENCES `template` (`idtemplate`) ON
DELETE CASCADE ON UPDATE CASCADE,

  CONSTRAINT `FK_dictionary_1` FOREIGN KEY (`idsubject`) REFERENCES `template` (`idtemplate`) ON
DELETE CASCADE ON UPDATE CASCADE

) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


--

-- Dumping data for table `dictionary`

--


LOCK TABLES `dictionary` WRITE;

/*!40000 ALTER TABLE `dictionary` DISABLE KEYS */;
```

```
INSERT INTO `dictionary` VALUES
(11,'is_a',1),(35,'is_a',1),(36,'is_a',1),(37,'is_a',1),(40,'is_a',1),(59,'is_a',1),(60,'is_a',1),(61
,'is_a',1),(66,'is_a',1),(67,'is_a',1),(68,'is_a',1),(69,'is_a',1),(70,'is_a',1),(77,'is_a',1),(4,'is
_a',2),(12,'is_a',2),(13,'is_a',2),(19,'is_a',2),(25,'is_a',2),(30,'is_a',2),(34,'is_a',2),(46,'is_a'
,2),(48,'is_a',2),(54,'is_a',2),(72,'is_a',2),(73,'is_a',2),(74,'is_a',2),(78,'is_a',2),(79,'is_a',2)
,(80,'is_a',2),(81,'is_a',2),(82,'is_a',2),(7,'is_a',3),(8,'is_a',3),(9,'is_a',3),(10,'is_a',3),(24,'
is_a',3),(28,'is_a',3),(31,'is_a',3),(32,'is_a',3),(33,'is_a',3),(41,'is_a',3),(47,'is_a',3),(49,'is_
a',3),(50,'is_a',3),(51,'is_a',3),(55,'is_a',3),(56,'is_a',3),(57,'is_a',3),(58,'is_a',3),(62,'is_a',
3),(64,'is_a',3),(4,'is_a',4),(14,'is_a',5),(23,'is_a',5),(26,'is_a',5),(38,'is_a',5),(42,'is_a',5),(
43,'is_a',5),(44,'is_a',5),(65,'is_a',5),(71,'is_a',5),(75,'is_a',5),(20,'is_a',6),(21,'is_a',6),(22,
'is_a',6),(27,'is_a',6),(29,'is_a',6),(39,'is_a',6),(45,'is_a',6),(52,'is_a',6),(53,'is_a',6),(63,'is
_a',6),(15,'is_a',78),(16,'is_a',78),(17,'is_a',79),(18,'is_a',79),(85,'is_part_of',84),(86,'is_part_
of',84),(88,'is_part_of',84),(1,'is_a',85),(2,'is_a',85),(3,'is_a',85),(4,'is_a',85),(5,'is_a',85),(6
,'is_a',85),(89,'is_a',88),(90,'is_a',88),(91,'is_a',88),(92,'is_part_of',84);

/*!40000 ALTER TABLE `dictionary` ENABLE KEYS */;

UNLOCK TABLES;


insert into template values (null,"Models");

insert into dictionary (idsubject,relationship,idobject)

SELECT t1.idtemplate,"is_a",t2.idtemplate FROM template t1

join template t2

where t1.name = "Models" and t2.name is null;


-- ----------------------------------------------------

DROP TABLE IF EXISTS `models`;

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` blob 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',

  PRIMARY KEY  (`idmodel`),

  UNIQUE KEY `Index_5` USING BTREE (`name`),

  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`),

  CONSTRAINT `FK_models_predicted` FOREIGN KEY (`predicted`) REFERENCES `template` (`idtemplate`) ON
UPDATE CASCADE,

  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_predictors` FOREIGN KEY (`predictors`) REFERENCES `template` (`idtemplate`)
ON UPDATE CASCADE

) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
```

```
-- -----------------------------------------------------
-- Table `tuples` for non-scalar values
-- -----------------------------------------------------
DROP TABLE IF EXISTS `tuples`;
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 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;

-- -----------------------------------------------------
-- Table `property_values` all values
-- -----------------------------------------------------
DROP TABLE IF EXISTS `property_values`;
CREATE TABLE  `property_values` (
  `id` int(10) unsigned NOT NULL auto_increment,
  `idproperty` int(10) unsigned NOT NULL,
  `idstructure` 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` USING BTREE (`idproperty`,`idstructure`),
  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` USING BTREE (`idproperty`,`idtype`),
  KEY `Index_8` (`idproperty`,`idvalue_string`),
  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
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;

-- -----------------------------------------------------
-- Table `quality_labels`
-- -----------------------------------------------------
DROP TABLE IF EXISTS `quality_labels`;
```

```
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  USING BTREE (`id`,`user_name`),
  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;


-- ----------------------------------------------------
-- Table `quality_pair`
-- ----------------------------------------------------
DROP TABLE IF EXISTS `quality_pair`;
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` USING BTREE (`idchemical`,`rel`),
  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  DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


-- ----------------------------------------------------
-- Table `quality_structure`
-- ----------------------------------------------------
DROP TABLE IF EXISTS `quality_chemicals`;
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` USING BTREE (`label`,`text`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
-- ------------------------------------------------------
-- Table `quality_structure`
-- ------------------------------------------------------
DROP TABLE IF EXISTS `quality_structure`;
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  USING BTREE (`idstructure`,`user_name`),
  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;


-- ------------------------------------------------------
-- Table `property_tuples` tuples of properties
-- ------------------------------------------------------
DROP TABLE IF EXISTS `property_tuples`;
CREATE TABLE  `property_tuples` (
  `idtuple` int(10) unsigned NOT NULL,
  `id` int(10) unsigned NOT NULL,
  PRIMARY KEY  USING BTREE (`idtuple`,`id`),
  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;


-- ------------------------------------------------------
-- Table `history`
-- ------------------------------------------------------
DROP TABLE IF EXISTS `history`;


CREATE TABLE  `history` (
  `version` int(11) NOT NULL auto_increment,
  `idstructure` int(11) unsigned NOT NULL,
  `structure` blob NOT NULL,
  `format` enum('SDF','CML','MOL') collate utf8_bin NOT NULL default 'CML',
```

```
  `label` enum('OK','UNKNOWN','ERROR') collate utf8_bin NOT NULL default 'UNKNOWN' COMMENT 'quality
label',
  `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',
  PRIMARY KEY  (`version`),
  KEY `idstructure` (`idstructure`),
  KEY `f_idstructure` (`idstructure`),
  KEY `FK_history_1` (`user_name`),
  CONSTRAINT `FK_history_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`) ON DELETE SET
NULL ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


-- ----------------------------------------------------
-- Table `src_dataset` datasets
-- ----------------------------------------------------
DROP TABLE IF EXISTS `src_dataset`;
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,
  PRIMARY KEY  (`id_srcdataset`),
  UNIQUE KEY `src_dataset_name` (`name`),
  KEY `FK_src_dataset_1` (`user_name`),
  KEY `FK_src_dataset_2` (`idreference`),
  CONSTRAINT `FK_src_dataset_2` FOREIGN KEY (`idreference`) REFERENCES `catalog_references`
(`idreference`) ON UPDATE CASCADE,
  CONSTRAINT `FK_src_dataset_1` FOREIGN KEY (`user_name`) REFERENCES `users` (`user_name`) ON DELETE
SET NULL ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


-- ----------------------------------------------------
-- Table `struc_dataset` structures per dataset
-- ----------------------------------------------------
DROP TABLE IF EXISTS `struc_dataset`;
CREATE TABLE IF NOT EXISTS  `struc_dataset` (
  `idstructure` int unsigned  NOT NULL ,
  `id_srcdataset` int 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
```

```
);

-- ----------------------------------------------------
-- Table `sessions` User sessions
-- ----------------------------------------------------
DROP TABLE IF EXISTS `sessions`;
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,
  `title` varchar(45) collate utf8_bin NOT NULL default 'Default',
  PRIMARY KEY  (`idsessions`),
  UNIQUE KEY `Index_3` USING BTREE (`title`,`user_name`),
  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 DEFAULT CHARSET=utf8 COLLATE=utf8_bin;

-- ----------------------------------------------------
-- Table `query` User queries per session
-- ----------------------------------------------------
DROP TABLE IF EXISTS `query`;
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` USING BTREE (`name`(255),`idsessions`),
  KEY `FK_query_1` (`idsessions`),
  KEY `FK_query_2` (`idtemplate`),
  CONSTRAINT `FK_query_2` FOREIGN KEY (`idtemplate`) REFERENCES `template` (`idtemplate`) ON DELETE
CASCADE ON UPDATE CASCADE,
  CONSTRAINT `FK_query_1` FOREIGN KEY (`idsessions`) REFERENCES `sessions` (`idsessions`) ON DELETE
CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin COMMENT='Structure queries';

-- ----------------------------------------------------
-- Table `query_results` Results of a user query
-- ----------------------------------------------------
DROP TABLE IF EXISTS `query_results`;
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` USING BTREE (`idquery`,`metric`),
  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;


-- ----------------------------------------------------
-- Table `funcgroups` Functional groups
-- ----------------------------------------------------
DROP TABLE IF EXISTS `funcgroups`;
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;


-- ----------------------------------------------------
-- Table `struc_fgroups` Functional groups per structure
-- ----------------------------------------------------
DROP TABLE IF EXISTS `struc_fgroups`;
CREATE TABLE  `struc_fgroups` (
  `idfuncgroup` int(10) unsigned NOT NULL,
  `idchemical` int(10) unsigned NOT NULL default '0',
  PRIMARY KEY  USING BTREE (`idfuncgroup`,`idchemical`),
  KEY `FK_struc_fgroups_1` (`idchemical`),
  CONSTRAINT `FK_struc_fgroups_2` FOREIGN KEY (`idfuncgroup`) REFERENCES `funcgroups` (`idfuncgroup`)
ON DELETE CASCADE ON UPDATE CASCADE,
  CONSTRAINT `FK_struc_fgroups_1` FOREIGN KEY (`idchemical`) REFERENCES `chemicals` (`idchemical`) ON
DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;


-- ----------------------------------------------------
-- Table `fp1024` 1024 length hashed fingerprints
-- ----------------------------------------------------
DROP TABLE IF EXISTS `fp1024`;
```

```
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`,`fp1
6`),
  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;


-- --------------------------------------------------------
-- Table `fp1024_struc` 1024 length hashed fingerprints
-- --------------------------------------------------------

DROP TABLE IF EXISTS `fp1024_struc`;
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  USING BTREE (`idchemical`,`idstructure`),
  KEY `fpall`
(`fp1`,`fp2`,`fp3`,`fp4`,`fp5`,`fp6`,`fp7`,`fp8`,`fp9`,`fp10`,`fp11`,`fp12`,`fp13`,`fp14`,`fp15`,`fp1
6`),
  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;


-- ------------------------------------------------------
-- Table `sk1024` 1024 structural keys
-- ------------------------------------------------------
DROP TABLE IF EXISTS `sk1024`;
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`,`fp1
6`),
  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;



-- --------------------------------------------------
-- Table `atom_distance`  Distance between atoms (discretized)
-- --------------------------------------------------
DROP TABLE IF EXISTS `atom_distance`;
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;


-- --------------------------------------------------
-- Table `atom_structure`  Distance between atoms (reference to atom_distance)
-- --------------------------------------------------
DROP TABLE IF EXISTS `atom_structure`;
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;


-- --------------------------------------------------
-- Table `version` Version
-- --------------------------------------------------
DROP TABLE IF EXISTS `version`;
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),
  PRIMARY KEY  (`idmajor`,`idminor`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 COLLATE=utf8_bin;
insert into version (idmajor,idminor,comment) values (2,10,"AMBIT2 schema");

-- ------------------------------------------------------
-- Sorts comma seperated strings
-- ------------------------------------------------------
DROP FUNCTION IF EXISTS `sortstring`;

DELIMITER $
CREATE FUNCTION `sortstring`(inString TEXT) RETURNS TEXT DETERMINISTIC
BEGIN
  DECLARE delim CHAR(1) DEFAULT ',';
  DECLARE strings INT DEFAULT 0;     -- number of substrings
  DECLARE forward INT DEFAULT 1;     -- index for traverse forward thru substrings
  DECLARE backward INT;   -- index for traverse backward thru substrings, position in calc.
substrings
  DECLARE remain TEXT;               -- work area for calc. no of substrings
  DECLARE swap1 TEXT;                 -- left substring to swap
  DECLARE swap2 TEXT;                 -- right substring to swap
  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 ;

-- ----------------------------------------------------
-- Generates sql given numerical and nominal property for a histogram of a numerical property
-- ----------------------------------------------------

DROP FUNCTION IF EXISTS `sql_xtab`;
DELIMITER $$

CREATE FUNCTION sql_xtab(property_num VARCHAR(128),property_nom VARCHAR(128), query INT,bins DOUBLE)
RETURNS TEXT 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 ;

-- ------------------------------------------------------
-- Generates sql given numerical and nominal property for a histogram of a numerical property
-- ------------------------------------------------------
DROP FUNCTION IF EXISTS `sql_dataset_xtab`;
DELIMITER $$

CREATE FUNCTION sql_dataset_xtab(property_num VARCHAR(128),property_nom VARCHAR(128), dataset
INT,bins DOUBLE) RETURNS TEXT
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 ;

DROP PROCEDURE IF EXISTS `p_dataset_xtab`;
DELIMITER $$

-- -------------------------------------------------------
-- Generates sql given numerical and nominal property for a histogram of a numerical property
-- -------------------------------------------------------
CREATE PROCEDURE p_dataset_xtab(IN property_num VARCHAR(128),property_nom VARCHAR(128),q INT,bins
DOUBLE)
LANGUAGE SQL
READS SQL DATA
CONTAINS SQL
SQL SECURITY DEFINER
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 ;
-- -------------------------------------------------------
-- Generates cross tab view  given numerical and nominal property
-- -------------------------------------------------------
```

```
DROP PROCEDURE IF EXISTS `p_xtab`;
DELIMITER $$

CREATE PROCEDURE p_xtab(IN property_num VARCHAR(128),property_nom VARCHAR(128),q INT,bins DOUBLE)
LANGUAGE SQL
READS SQL DATA
CONTAINS SQL
SQL SECURITY DEFINER
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 ;
-- ------------------------------------------------------
-- numeric property values
-- ------------------------------------------------------
DROP VIEW IF EXISTS `values_number`;
create view `values_number` as
SELECT id,idproperty,idstructure,value_num as value,status,user_name
FROM property_values where value_num is not null;

-- ------------------------------------------------------
-- string property values
-- ------------------------------------------------------

DROP VIEW IF EXISTS `values_string`;
create view `values_string` as
SELECT id,idproperty,idstructure,if(status="TRUNCATED",text,value) as value,status,user_name,name
FROM properties join property_values using(idproperty) join property_string using(idvalue_string)
where idvalue_string is not null;

-- ------------------------------------------------------
-- all property values
-- ------------------------------------------------------
DROP VIEW IF EXISTS `values_all`;
create view values_all as
SELECT idstructure,idproperty,name,null as value_string,value_num as value_number,idreference FROM
properties join property_values using(idproperty)
where value_num is not null
union
SELECT idstructure,idproperty,name,value as value_string,null as value_number,idreference FROM
properties join property_values using(idproperty) join property_string using(idvalue_string)
where idvalue_string is not null;

-- ------------------------------------------------------
```

```
-- ontology
-- -----------------------------------------------------
DROP VIEW IF EXISTS `ontology`;
create view ontology as
SELECT t1.idtemplate as subjectid,t2.idtemplate as objectid,t1.name as subject,relationship,t2.name
as object FROM template as t1 join dictionary as d on t1.idtemplate=d.idsubject join template as t2
on d.idobject=t2.idtemplate;


-- -----------------------------------------------------
-- Template definitions
-- -----------------------------------------------------
DROP VIEW IF EXISTS `template_properties`;
create view template_properties as
SELECT idtemplate,template.name as template,idproperty,properties.name as property FROM template join
template_def using(idtemplate) join properties using(idproperty);


-- -----------------------------------------------------
-- default users (<removed>, <removed>)
-- -----------------------------------------------------


-- insert into roles (role_name) values ("ambit_guest");
-- insert into roles (role_name) values ("ambit_admin");
-- insert into roles (role_name) values ("quality");
-- insert into users
(user_name,password,email,lastname,registration_date,registration_status,keywords,webpage) values
("<removed>","<removed>","<removed>","<removed>",now(),"confirmed","<removed>","<removed>");
-- insert into users
(user_name,password,email,lastname,registration_date,registration_status,keywords,webpage) values
("<removed>","<removed>","<removed>","<removed>",now(),"confirmed","<removed>","<removed>");
-- insert into user_roles (user_name,role_name) values ("<removed>","ambit_guest");
-- insert into user_roles (user_name,role_name) values ("<removed>","ambit_admin");


-- -----------------------------------------------------
-- default dictionary entries
-- -----------------------------------------------------
insert into catalog_references (idreference,title,url) values (1,"CAS Registry
Number","http://www.cas.org");
-- insert into dictionary (iddictionary,idparent,name,idreference) values (null,0,"CAS",1);


insert into catalog_references (idreference,title,url) values (2,"IUPAC
name","http://www.iupac.org");
-- insert into dictionary (iddictionary,idparent,name,idreference) values (null,0,"IUPAC Name",2);



-- Grants for ambit2@localhost
-- REVOKE ALL PRIVILEGES ON ambit2.* FROM '<removed>'@'localhost';
-- REVOKE ALL PRIVILEGES ON ambit2.* FROM '<removed>'@'localhost';
-- GRANT USAGE ON ambit2.* TO '<removed>'@'localhost' IDENTIFIED BY PASSWORD '<removed>';
-- GRANT ALL PRIVILEGES ON ambit2.* TO '<removed>'@'localhost' WITH GRANT OPTION;
-- GRANT SELECT, INSERT, UPDATE, DELETE, SHOW VIEW ON ambit2.* TO '<removed>'@'localhost' IDENTIFIED
BY PASSWORD '<removed>';
```

```
-- GRANT EXECUTE ON FUNCTION sortstring TO '<removed>'@'localhost'
--- GRANT CREATE TEMPORARY TABLEs on ambit2.* to '<removed>'@'%'

SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;
```

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"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:otee="http://www.opentox.org/echaEndpoints.owl#"
    xmlns="http://purl.org/dc/elements/1.1/"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:swrl="http://www.w3.org/2003/11/swrl#"
    xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.opentox.org/echaEndpoints.owl">
  <owl:Ontology rdf:about="">
    <owl:imports rdf:resource="http://protege.stanford.edu/plugins/owl/dc/protege-dc.owl"/>
  </owl:Ontology>
  <owl:Ontology rdf:about="#">
    <owl:imports rdf:resource="http://protege.stanford.edu/plugins/owl/dc/protege-dc.owl"/>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >ECHA endpoints</rdfs:comment>
  </owl:Ontology>
  <owl:Class rdf:about="#Dissociation_constant_(pKa)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Dissociation constant (pKa)</title>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="PhysicoChemicalEffects"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Vapour_pressure">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PhysicoChemicalEffects"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Vapour pressure</title>
  </owl:Class>
  <owl:Class rdf:ID="Gastrointestinal_absorption">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="ToxicoKinetics"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Gastrointestinal absorption</title>
  </owl:Class>
  <owl:Class rdf:ID="EcotoxicEffects">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Endpoints"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Ecotoxic effects</title>
  </owl:Class>
  <owl:Class rdf:ID="Carcinogenicity">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="HumanHealthEffects"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Carcinogenicity</title>
  </owl:Class>
  <owl:Class rdf:about="#Eye_irritation/corrosion">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#HumanHealthEffects"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Eye irritation/corrosion</title>
  </owl:Class>
  <owl:Class rdf:about="#Long-term_toxicity_to_Daphnia_(lethality,_inhibition_of_reproduction)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
```

```
  >Long-term toxicity to Daphnia (lethality, inhibition of reproduction)</title>
  <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
</owl:Class>
<owl:Class rdf:ID="Photocarcinogenicity">
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Photocarcinogenicity</title>
  <rdfs:subClassOf>
    <owl:Class rdf:about="#HumanHealthEffects"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="AcuteInhalationToxicity">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#HumanHealthEffects"/>
  </rdfs:subClassOf>
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Acute inhalation toicity</title>
</owl:Class>
<owl:Class rdf:about="#Short-term_toxicity_to_algae_(inhibition_of_the_exponential_growth_rate)">
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Short-term toxicity to algae (inhibition of the exponential growth rate)</title>
  <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
</owl:Class>
<owl:Class rdf:ID="Toxicity_to_birds">
  <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Toxicity to birds</title>
</owl:Class>
<owl:Class rdf:ID="Photomutagenicity">
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Photomutagenicity</title>
  <rdfs:subClassOf>
    <owl:Class rdf:about="#HumanHealthEffects"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="#Air-_water_partition_coefficient_(Henry&apos;s_law_constant,_H)">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#PhysicoChemicalEffects"/>
  </rdfs:subClassOf>
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Air- water partition coefficient (Henry`s law constant, H)</title>
</owl:Class>
<owl:Class rdf:ID="Water_solubility">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#PhysicoChemicalEffects"/>
  </rdfs:subClassOf>
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Water solubility</title>
</owl:Class>
<owl:Class rdf:ID="Protein-binding">
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Protein-binding</title>
  <rdfs:subClassOf>
    <owl:Class rdf:about="#ToxicoKinetics"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Vegetation-air_partition_coefficient">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="EnvironmentalFateParameters"/>
  </rdfs:subClassOf>
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Vegetation-air partition coefficient</title>
</owl:Class>
<owl:Class rdf:about="#HumanHealthEffects">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#Endpoints"/>
  </rdfs:subClassOf>
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Human health effects</title>
</owl:Class>
<owl:Class rdf:about="#Toxicity_to_soil_invertebrates_(survival,_growth,_reproduction)">
  <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Toxicity to soil invertebrates (survival, growth, reproduction)</title>
  <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
</owl:Class>
```

```
  <owl:Class rdf:ID="Vegetation-soil_partition_coefficient">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Vegetation-soil partition coefficient</title>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#EnvironmentalFateParameters"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#Adsorption/Desorption_in_soil">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#EnvironmentalFateParameters"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Adsorption/Desorption in soil </title>
  </owl:Class>
  <owl:Class rdf:about="#EnvironmentalFateParameters">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Endpoints"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Environmental fate parameters </title>
  </owl:Class>
  <owl:Class rdf:ID="s45">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Ocular membrane penetration</title>
  </owl:Class>
  <owl:Class rdf:about="#Organic_carbon-sorption_partition_coefficient_(organic_carbon;_Koc)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Organic carbon-sorption partition coefficient (organic carbon; Koc)</title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class rdf:about="#In_vivo_pre-,_peri-
,_post_natal_development_and_/_or_fertility_(1_or_2_gen._Study_or_enhanced_1_gen_study)">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >In vivo pre-, peri-, post natal development and / or fertility (1 or 2 gen. Study or enhanced 1
gen study) </title>
  </owl:Class>
  <owl:Class rdf:about="#Adsorption/Desorption_in_sediment">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Adsorption/Desorption in sediment</title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class
rdf:about="#In_vitro_reproductive_toxicity_(e.g._embryotoxic_effects_in_cell_culture_such_as_embryo_s
tem_cells)">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >In vitro reproductive toxicity (e.g. embryotoxic effects in cell culture such as embryo stem
cells)  </title>
  </owl:Class>
  <owl:Class rdf:about="#Toxicity_to_plants_(leaves,_seed_germination,_root_elongation)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Toxicity to plants (leaves, seed germination, root elongation)</title>
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="Surface_tension">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PhysicoChemicalEffects"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Surface tension</title>
  </owl:Class>
  <owl:Class rdf:ID="AcuteOralToxicity">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Acute oral toxicity</title>
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="Blood-lung_barrier_penetration">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
```

```
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Blood-lung barrier penetration</title>
  </owl:Class>
  <owl:Class rdf:ID="Repeated_dose_toxicity">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Repeated dose toxicity </title>
  </owl:Class>
  <owl:Class rdf:ID="DNA-binding">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >DNA-binding</title>
  </owl:Class>
  <owl:Class
rdf:about="#Other_(e.g._inhibition_of_specific_enzymes_involved_in_hormone_synthesis_or_regulation,_specify_enzyme(s)_and_hormone)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Other (e.g. inhibition of specific enzymes involved in hormone synthesis or regulation, specify
enzyme(s) and hormone)</title>
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="Vegetation-water_partition_coefficient">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Vegetation-water partition coefficient</title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class rdf:ID="Photosensitisation">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Photosensitisation</title>
  </owl:Class>
  <owl:Class rdf:ID="Boiling_point">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#PhysicoChemicalEffects"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Boiling point</title>
  </owl:Class>
  <owl:Class rdf:about="#PhysicoChemicalEffects">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Physicochemical effects </title>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Endpoints"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#Octanol-water_distribution_coefficient_(D)">
    <rdfs:subClassOf rdf:resource="#PhysicoChemicalEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Octanol-water distribution coefficient (D)</title>
  </owl:Class>
  <owl:Class rdf:ID="SkinSensitisation">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Skin sensitisation</title>
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="PersistenceAbioticDegradationAir">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Persistence"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Persistence: Abiotic degradation in air (Phototransformation)</title>
  </owl:Class>
  <owl:Class rdf:about="#Octanol-water_partition_coefficient_(Kow)">
    <rdfs:subClassOf rdf:resource="#PhysicoChemicalEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Octanol-water partition coefficient (Kow)</title>
  </owl:Class>
  <owl:Class rdf:about="#Ready/not_ready_biodegradability">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Ready/not ready biodegradability</title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
```

```
  <owl:Class rdf:ID="BCF_fish">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="BioConcentration"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >BCF fish</title>
  </owl:Class>
  <owl:Class rdf:ID="s39">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Metabolism (including metabolic clearance)</title>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Hydrolysis">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Hydrolysis </title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class rdf:ID="Mutagenicity">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Mutagenicity </title>
  </owl:Class>
  <owl:Class rdf:about="#Short-term_toxicity_to_Daphnia_(immobilisation)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Short-term toxicity to Daphnia (immobilisation)</title>
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  </owl:Class>
  <owl:Class rdf:about="#Receptor-binding_(specify_receptor)">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Receptor-binding (specify receptor)</title>
  </owl:Class>
  <owl:Class rdf:about="#Persistence">
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class rdf:ID="Melting_point">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Melting point</title>
    <rdfs:subClassOf rdf:resource="#PhysicoChemicalEffects"/>
  </owl:Class>
  <owl:Class rdf:about="#Octanol-air_partition_coefficient_(Koa)">
    <rdfs:subClassOf rdf:resource="#PhysicoChemicalEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Octanol-air partition coefficient (Koa)</title>
  </owl:Class>
  <owl:Class rdf:about="#Endpoints">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Endpoints</title>
  </owl:Class>
  <owl:Class rdf:about="#Toxicity_to_soil_microorganisms_(inhibition_of_C-
mineralisation,_inhibition_of_N-mineralisation,_other)">
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Toxicity to soil microorganisms (inhibition of C-mineralisation, inhibition of N-mineralisation,
other)</title>
  </owl:Class>
  <owl:Class rdf:about="#Toxicity_to_earthworms_(survival,_growth,_reproduction)">
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Toxicity to earthworms (survival, growth, reproduction)</title>
  </owl:Class>
  <owl:Class rdf:ID="Blood-brain_barrier_penetration">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Blood-brain barrier penetration</title>
  </owl:Class>
  <owl:Class rdf:ID="PersistenceAbioticDegradationWater">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Persistence: Abiotic degradation in water</title>
    <rdfs:subClassOf rdf:resource="#Persistence"/>
```

```
  </owl:Class>
  <owl:Class rdf:ID="BioAccumulation">
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Bioaccumulation</title>
  </owl:Class>
  <owl:Class
rdf:about="#Microbial_inhibition_(activated_sludge_respiration_inhibition,_inhibition_of_nitrificatio
n,_other)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Microbial inhibition (activated sludge respiration inhibition, inhibition of nitrification,
other)</title>
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  </owl:Class>
  <owl:Class rdf:about="#Long-
term_toxicity_to_fish_(egg/sac_fry,_growth_inhibition_of_juvenile_fish,_early_life_stage,_full_life_c
ycle)">
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Long-term toxicity to fish (egg/sac fry, growth inhibition of juvenile fish, early life stage,
full life cycle)</title>
  </owl:Class>
  <owl:Class rdf:about="#BioConcentration">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Bioconcentration </title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class rdf:ID="Direct_photolysis">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Direct photolysis</title>
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
  </owl:Class>
  <owl:Class rdf:about="#Biodegradation_time_frame_(primary,_ultimate_degradation)">
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Biodegradation time frame (primary, ultimate degradation)</title>
  </owl:Class>
  <owl:Class rdf:ID="Blood-testis_barrier_penetration">
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Blood-testis barrier penetration</title>
  </owl:Class>
  <owl:Class rdf:ID="Respiratory_sensitisation">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Respiratory sensitisation</title>
  </owl:Class>
  <owl:Class rdf:ID="Skin_penetration">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Skin penetration</title>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#Long-term_toxicity_(survival,_growth,_reproduction)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Long-term toxicity (survival, growth, reproduction)</title>
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="Placental_barrier_penetration">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Placental barrier penetration</title>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#ToxicoKinetics"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:about="#ToxicoKinetics">
    <rdfs:subClassOf rdf:resource="#Endpoints"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Toxicokinetics </title>
  </owl:Class>
  <owl:Class rdf:ID="PersistenceBiodegradation">
```

```
    <rdfs:subClassOf rdf:resource="#Persistence"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Persistence: Biodegradation</title>
  </owl:Class>
  <owl:Class rdf:ID="BAF_other_organisms">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >BAF other organisms </title>
    <rdfs:subClassOf rdf:resource="#BioAccumulation"/>
  </owl:Class>
  <owl:Class rdf:about="#Short_term_toxicity_(feeding,_gavage,_other)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Short term toxicity (feeding, gavage, other)</title>
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="Oxidation">
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Oxidation </title>
  </owl:Class>
  <owl:Class rdf:about="#Receptor_binding_and_gene_expression_(specify_receptor)">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Receptor binding and gene expression (specify receptor)</title>
  </owl:Class>
  <owl:Class rdf:about="#Toxicity_to_sediment_organisms_(survival,_growth,_reproduction)">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Toxicity to sediment organisms (survival, growth, reproduction)</title>
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="BAF_fish">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >BAF fish</title>
    <rdfs:subClassOf rdf:resource="#BioAccumulation"/>
  </owl:Class>
  <owl:Class rdf:ID="Acute_photoirritation">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Acute photoirritation</title>
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
  </owl:Class>
  <owl:Class rdf:ID="In_vivo_pre-natal-developmental_toxicity">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >In vivo pre-natal-developmental toxicity</title>
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
  </owl:Class>
  <owl:Class rdf:about="#Indirect_photolysis_(OH-radical_reaction,_ozone-radical_reaction,_other)">
    <rdfs:subClassOf rdf:resource="#EnvironmentalFateParameters"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)</title>
  </owl:Class>
  <owl:Class rdf:ID="AcuteDermalToxicity">
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Acute dermal toxicity</title>
  </owl:Class>
  <owl:Class rdf:about="#Acute_toxicity_to_fish_(lethality)">
    <rdfs:subClassOf rdf:resource="#EcotoxicEffects"/>
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Acute toxicity to fish (lethality)</title>
  </owl:Class>
  <owl:Class rdf:ID="BCF_other_organisms">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >BCF other organisms </title>
    <rdfs:subClassOf rdf:resource="#BioConcentration"/>
  </owl:Class>
  <owl:Class rdf:ID="SkinIrritationCorrosion">
    <title rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Skin irritation /corrosion</title>
    <rdfs:subClassOf rdf:resource="#HumanHealthEffects"/>
  </owl:Class>
  <otee:SkinSensitisation rdf:about="http://purl.org/dc/elements/1.1/SkinSensitisation"/>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.1, Build 536)  http://protege.stanford.edu -->
```