

OpenTox Predictive Toxicology Framework: toxicological ontology and semantic media wiki-based OpenToxipedia

Olga Tcheremenskaia* (A), Romualdo Benigni (A), Ivelina Nikolova (B), Nina Jeliaskova (B), Sylvia E. Escher (C), Helvi Grimm (C), Thomas Baier (C), Vladimir Poroikov (D), Alexey Lagunin (D), Micha Rautenberg (E) and Barry Hardy* (F)

(A) Istituto Superiore di Sanità, Environment and Health Department, Viale Regina Elena 299, Rome 00161, Italy; (B) Ideaconsult Ltd, A. Kanchev 4, Sofia 1000, Bulgaria; (C) Fraunhofer Institute for Toxicology & Experimental Medicine, Nikolai-Fuchs-Str. 1, 30625 Hannover, Germany; (D) Institute of Biomedical Chemistry of Russian Academy of Sciences, Pogodinskaya street 10, 119121 Moscow, Russia; (E) In silico Toxicology, Altkircher Str. 4, CH-4052 Basel, Switzerland; (F) Douglas Connect, Baermeggenweg 14, CH-4314 Zeiningen, Switzerland

ABSTRACT

The OpenTox Framework, developed by the partners in the EC FP7 OpenTox project, aims at providing a unified access to toxicity data, predictive models and validation procedures (B. Hardy, 2010). Interoperability of resources is achieved using a common information model, based on an OpenTox OWL-DL ontology and related ontologies, describing predictive algorithms, models and toxicity data. As toxicological data may come from different, heterogeneous sources, a deployed ontology unifying the terminology and the resources is critical for the rational and reliable organization of the data, and its automatic processing. Up to now the following related ontologies have been developed for OpenTox: Toxicological ontology – listing the toxicological endpoints; Organ system ontology – addressing targets/examinations and organs observed in *in vivo* studies; ToxML ontology – representing semi-automatic conversion of the ToxML schema; ToxLink – ToxCast assays ontology; OpenTox ontology – representation of OpenTox framework components: chemical compounds, datasets, algorithms, models and validation web services; Algorithms Ontology – types of algorithms. Besides being defined in an ontology, OpenTox components are made available through standardized REST web services, where every compound, data set or predictive method has a unique resolvable address (URI), used to retrieve its Resource Description Framework (RDF) representation, or to initiate the associated calculations and generate new RDF-based resources. The services support the integration of toxicity and chemical data from various sources, the generation and validation of computer models for toxic effects, seamless integration of new algorithms and scientifically sound validation routines and provide a flexible framework, which allow building arbitrary number of applications, tailored to solving different problems by end users (e.g. toxicologists).

* To whom correspondence should be addressed. Email: Barry.Hardy -(at)-douglasconnect.com

1 INTRODUCTION

OpenTox (OT) was funded by FP7 to develop a framework for predictive toxicology modelling and application development. Based on the framework of web services two initial OT web-applications have been made available: ToxPredict¹ that predicts the activity of a chemical structure submitted by the user in respect to a given toxicity endpoint, and ToxCreate² that creates a predictive toxicology model from a user-submitted dataset.

Ontology definition is important for OT, as information can be integrated in a more efficient and reliable manner, thus reducing the cost, maintenance and risk of application development and deployment. At the moment, our toxicological ontology structure aims to cover five “critical” toxicity study types: carcinogenicity, *in vitro* and *in vivo* mutagenicity from micronucleus assays, repeated dose toxicity and aquatic toxicity studies. Even though several ontologies for the biomedical field are publicly available, currently a systematic ontology for toxicological effects and predictive toxicology is not covered by the OBO Foundry³ or Biportal ontology depositories⁴. Whenever possible, we are trying to integrate relevant information of neighboring ontologies, such as the Foundational Model of Anatomy (FMA), Ontology for Biomedical Investigation (OBI), NCI Thesaurus, and SNOMED Clinical Terms together with the ToxML (Toxicology XML standard) schema⁵. The Organs Ontology developed is very closely linked to the INHAND initiative (International Harmonization of Nomenclature and Diagnostic Criteria for Lesions in Rats and Mice). INHAND aims to develop for the first time an internationally accepted

¹ ToxPredict <http://www.toxpredict.org/>

² ToxCreate <http://www.toxcreate.org/>

³ The OBO Foundry <http://www.obofoundry.org/>

⁴ NCBO BioPortal <http://biportal.bioontology.org/>

⁵ Leadscope ToxML Schema www.leadscope.com/toxml.php

standardized vocabulary for neoplastic and non-neoplastic lesions as well as the definition of diagnostic features for organ systems observed in *in vivo* studies. Recently, the description of the respiratory system has been published (R. Renne, 2009).

OpenToxipedia⁶ is a new related community resource of toxicology terminology organized by means of Semantic Media Wiki. OpenToxipedia allows creating, adding, editing and keeping terms used in both experimental toxicology and *in silico* toxicology. The particular importance of OpenToxipedia relies on the description of all the terms used in OT applications such as ToxPredict and ToxCreate.

2 METHODS

The construction of formal ontology follows relatively established principles in knowledge representation. We have taken into consideration those available for biomedical ontology development, particularly the OBO Foundry principles. An open, public approach to ontology development supports current and future collaborations with different projects. We use the DL species of the Web Ontology Language (OWL DL) supported by the Protégé OWL editor. An overview of the OT ontology is given on the public area of the OT website⁷ together with instructions on how to enter the OT Collaborative Protégé Server and contribute to existing OWL projects.

Some of the ontologies are manually created from scratch, others partially reuse existing ones and extend them with task related concepts and relations. The ToxML ontology is semi-automatically generated from the existing ToxML schema by parsing it to OWL and applying specific rules, which convey the semantics and remove redundant information in the new format.

OpenToxipedia has been developed using the Semantic Media Wiki (SMW). It was created manually by experts in the fields of *in silico* and experimental toxicology on the basis of known regulatory documents, glossaries, dictionaries and some primary publications. All registered members are welcome to add new entries, suggest definitions and edit the existing resource at www.opentoxipedia.org. The OpenToxipedia is curated by OT toxicology experts.

SMW was chosen for the OpenToxipedia representation for the following main reasons: it enables automatic processing of the wiki knowledge base; it gives a possibility for data transfer between RDF and SMW through SPARQL. SMW will facilitate the automatic data exchange between OpenToxipedia, the ontologies and their use by the OpenTox web services dealing with RDF data. The SMW is a collaborative system, supports versioning, RDF export,

tools to lock pages by a curator (fixing a validated vocabulary) and the possibility to add annotation without changing the ontology/rdf information.

3 RESULTS

Up to now, six ontologies have been made available through the OT Collaborative Protégé Server:

- Toxicological ontology;
- Organ system ontology;
- ToxML ontology;
- OpenTox ontology, representing components of OpenTox web services framework;
- Algorithm types ontology;
- ToxLink (ToxCast assays ontology).

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

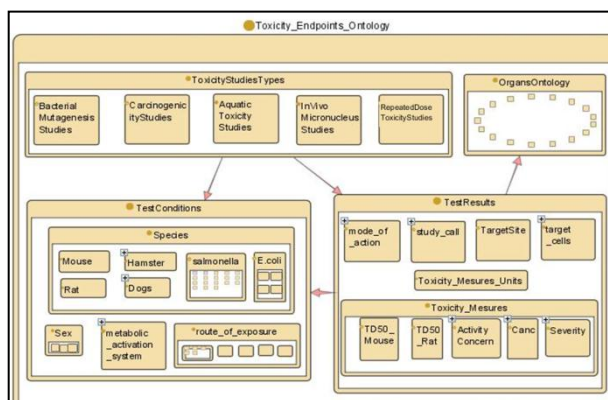


Fig. 1. OT toxicological ontology structure.

The “target sites” toxicological class is to be linked to the Organ system ontology, developed by the Fraunhofer Institute for Toxicology & Experimental Medicine. The Organs Ontology is one of the most challenging ontology classes addressing targets, examinations and organs observed in *in vivo* studies such as repeated dose toxicity and carcinogenicity. The ontology includes the detailed description of organs starting from organs systems down to histological components. It was decided to use a hierarchical structure starting

⁶ OpenToxipedia www.opentoxipedia.org

⁷ OpenTox ontology development page www.opentox.org/dev/ontology

with the organs system (e.g. digestive system) instead of orientating the ontology on the examinations performed in guideline studies such as histopathology, necropsy, and clinical observations. So the principal structure of the organs ontology is as follows:

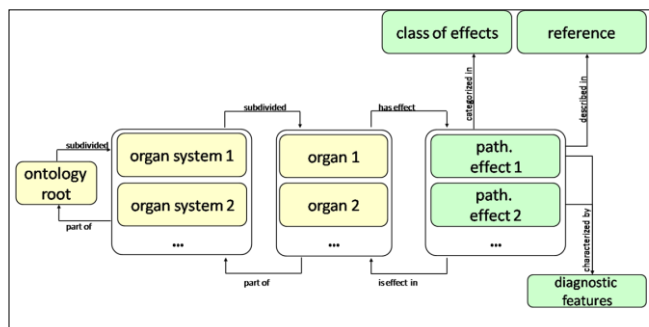
- Class Organs system - Subclass Organs system
- |- Class Target organs - Subclass Target organs 1 to N
- |- Class Histopathology - Subclasses if needed

At the moment the Organs Ontology includes 12 organs systems: digestive system, respiratory system, circulatory system, endocrine system, male genital system, female genital system, hematopoietic system, integumentary system, nervous system and special sense organs, urinary system, musculoskeletal system, immune system and lymphatic organs. Synonyms are included to account for differences in terminologies. It focuses on the organs observed in rodents, which are frequently used for toxicity testing. Species specificity will be introduced, when combining the organ ontology with the toxicological endpoint ontology.

Currently, the Toxicological Effects Ontology comprises neoplastic and non-neoplastic effects observed in repeated dose and cancer studies. Endpoint specificity of the effects will be included when combining the organ/effect root ontology with the toxicological endpoint ontology. The effects ontology consists of three main parts: classes of effects, linked to pathological effects, which are further linked to detailed diagnostic features as agreed in the INHAND initiative. Its functionality has been initially developed for the respiratory tract. The structure of the combined organ and effects ontology is depicted in Figure 2.

Fig. 2. Overview of the structure of the combined organ (in orange) and effect (in green) ontologies

The ToxML ontology is a semi-automatic conversion of



ToxML schema to OWL-DL. The most recent ToxML release has a comprehensive, well-structured scheme for many toxicity studies (carcinogenicity, *in vitro* mutagenicity, *in vivo* micronucleus, repeated dose toxicity) which fit well the OpenTox purposes. This was verified by manually mapping various existing database entries to the ToxML schema. The resulting ontology will be applied as a media to reference/annotate the contents of databases coming from various sources and toxicity studies. Our purpose is not only to develop a cross database matching schema but also to bene-

fit from the powerful reasoning mechanism that OWL offers to inference on existing facts in the databases. In order to use ToxML as a scheme for accommodation of our data we need to overcome the issues raised by the nature of the XML description: there exist many fields with free text instead of named concepts; standardized vocabularies for many classes do not exist (e.g. target sites, mode of action, route of exposure), therefore some classes and properties are named by more than one label and others have labels which are ambiguous; the XML nested structure does not follow the natural IS-A relation used for subclassing in OWL. For this reason *ad hoc* rules for conversion are implemented. The resulting ontology has a flat structure representing numerous relations except the IS-A relation, since IS-A does not apply to the concepts in use. Ambiguous labels are unified and a step towards label standardization is achieved, where possible object type properties are introduced instead of data type ones; thus the referenced values remain named instead of string values.

The OpenTox ontology⁸ provides a common information model for the most common components, found in any application, providing predictive toxicology functionality, namely chemical compounds, datasets of chemical compounds, data processing algorithms, machine learning algorithms, predictive models and validation routines. The OpenTox framework exposes REST web services, corresponding to each of these common components. A generic OWL representation is defined for every component (e.g. every OT dataset is a subclass of `ot:Dataset`, every algorithm is subclass of `ot:Algorithm` and every model is a subclass of `ot:Model`). This allows unified representation across diverse data and algorithms, and a uniform interface to data processing services, which take generic `ot:Dataset` resources on input and generate generic `ot:Dataset` resources on output. Specific types of algorithms are described in the algorithm types ontology and even more details of descriptor calculation algorithms are specified via the Blue Obelisk ontology (Guha at al., 2006) of cheminformatics algorithms (e.g. algorithm references, descriptor categories) and extensions, specifically developed to cover algorithms developed by OpenTox developers. Assigning specific information about the datasets, properties and types of algorithms and models is done via linking to the relevant ontologies, for example by subclass-ing (`rdf:type`), `owl:sameAs` links, or Blue Obelisk ontology `bo:instanceOf` predicate.

The simultaneous use of OT datasets and compound properties as resources of generic `ot:Dataset` type and `ot:Feature` type in the OT ontology, and linking to specific toxicology ontologies, provides a flexible mechanism for annotation. It allows users of OT web services to upload datasets of chem-

⁸ OpenTox Ontology <http://www.opentox.org/api/1.1/opentox.owl>

ical compounds and arbitrary named properties of the compounds. The datasets are converted into a uniform owl:Dataset representation and chemical compound properties can be manually annotated with the proper terms from toxicology ontologies. The annotation and assigning of owl:sameAs links is currently only done manually, via OT REST web service interface, which modifies the relevant resource representation by adding/modifying triples. In principle, more sophisticated techniques could be applied, and the corresponding RDF representation updated via the same REST interface. This approach is currently used to enter and represent data in OT services and applications. Description of one of the OT API implementations, and examples of RDF representation of various resources is provided in (N. Jeliaskova, V. Jeliaskov, 2011).

The sixth ontology project initiated is the ToxLink ontology representing the ToxCast assays from the US EPA. This development is a collaborative effort of OpenTox with ToxCast⁹ to provide an ontological description of *in vitro* toxicological assays.

At present, OpenToxipedia contains 862 toxicological terms with description and literature references classified into 26 categories (see Figure 3).

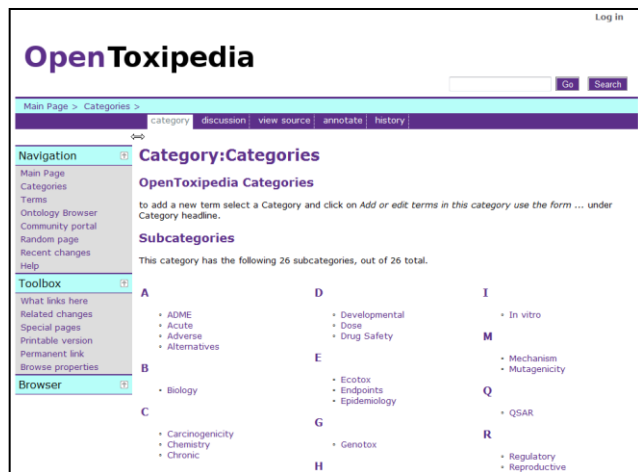


Fig. 3. OpenToxipedia categories for predictive toxicology.

The terms can be browsed either by category or in alphabetical order. Specialists in different toxicology fields are invited to take part in the creation and curation of OpenToxipedia. It can be used as a compendium for free available predictive toxicology resources supporting the application and development of the standards for representation of toxicology data, vocabulary and ontology development needed by OpenTox use cases and web services. The following rules for term management in OpenToxipedia have been developed: (i) Add terms – any registered user (curators receive a

message and decide what additions are approved and will become publicly available); (ii) Edit description of terms – curators; (iii) Add remarks - any registered user (curators receive an alert message).

4 DISCUSSION AND CONCLUSIONS

The need for speeding up the toxicological assessment of chemicals, and of using less animals and more inexpensive tools has strongly stimulated the development of predictive toxicology and of structure-based approaches. A wide spectrum of predictive approaches applied to toxicity exist today, including read-across, regulatory categories, and (Quantitative) structure-activity relationship ((Q)SAR) modelling. All these predictive approaches share the need of a highly structured information as a starting point: the definition of ontology and of controlled vocabulary is a crucial requirement in order to standardize and organize the chemical and toxicological data on which the predictive toxicology methods build on. In addition, the availability of ontology specific for predictive toxicology is crucial to the interoperability of OpenTox services and other platforms and software in developing and deploying user applications. The ontology will be submitted to the Biportal website for dissemination and feedback.

ACKNOWLEDGEMENTS

OpenTox - An Open Source Predictive Toxicology Framework, www.opentox.org, is funded under the EU Seventh Framework Program: HEALTH-2007-1.3-3 Promotion, development, validation, acceptance and implementation of QSARs (Quantitative Structure-Activity Relationships) for toxicology, Project Reference Number Health-F5-2008-200787 (2008-2011).

Project Partners

Douglas Connect, In Silico Toxicology, Ideaconsult, Istituto Superiore di Sanita', Technical University of Munich, Albert Ludwigs University Freiburg, National Technical University of Athens, David Gallagher, Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Seascape Learning and, The Fraunhofer Institute for Toxicology & Experimental Medicine.

REFERENCES

- B. Hardy, N. Douglas et al. (2010), Collaborative Development of Predictive Toxicology Applications, *Journal of Cheminformatics*, 2:7; doi:10.1186/1758-2946-2-7.
- R. Renne, A. Brix et al. (2009), Proliferative and Nonproliferative Lesions of the Rat and Mouse Respiratory Tract, *Toxicologic Pathology*, 37: 5-73
- R. Guha, M.T. Howard et al. (2006), The Blue Obelisk- interoperability in chemical informatics, *May-Jun*; 46(3): 991-8.
- N. Jeliaskova, V. Jeliaskov (2011), AMBIT RESTful web services: an implementation of the OpenTox application programming interface, *Journal of Cheminformatics*, 3:18; doi:10.1186/1758-2946-3-18

⁹ ToxCast www.epa.gov/ncct/toxcast/