

# OpenTox: an open source platform for toxicity prediction

**OpenTox Meeting, Munich, August 2011**

David Gallagher <david@cacheresearch.com>

Sunil Chawla <sunil@seascapelearning.com>

Barry Hardy <barry.hardy@douglasconnect.com> (P.I.)

# Abstract

The new European Union (EU) REACH chemical legislation will require over 9 million additional test animals, if no alternative methods for toxicity prediction are accepted. However, the number of test animals could be significantly reduced by utilizing existing experimental data in conjunction with (Quantitative) Structure Activity Relationship ((Q)SAR) models.

To address the challenge, the European Commission has funded the OpenTox ([www.OpenTox.org](http://www.OpenTox.org)) project to develop an open source web-service-based framework, that provides unified access to experimental toxicity data, in Silico models (including (Q)SAR), and validation/reporting procedures.

Now, in the final year of the initial three-year project, the current state of architecture, Open API, algorithms, ontologies, and approach to web services will be presented. Our experiences on current collaborative approaches aiming to combine OpenTox with other systems such as CERF, Bioclipse, CDK, and SYNERGY to create “superinteroperable K-infrastructure” will be discussed both in terms of conceptual promise and implementation reality.

# Topics

- **New EU legislation “REACH”** *David Gallagher*
- **The chemistry challenges**
- **Prototype applications (Tox prediction)**
- **User Needs** *Sunil Chawla*
- **Semantic Web for Predictive Toxicology**
- **Key Components**
- **Web Service Interoperability**
- **Achievements**

*Barry Hardy (P.I.)*

# Background - REACH

**“Registration, Evaluation, Authorisation & Restriction of Chemicals”**

**European Union Legislation, 2007**

***“...protecting human health and the environment...”*** Geert Dancet, ECHA

All chemicals imported or manufactured in Europe must be registered with European Chemicals Agency (ECHA)

phased threshold: Dec. 2010 > 1,000 tons p.a., 2018 > 1 ton p.a.

**Responsibility:**

**Manufacturers and importers to  
provide safety information & manage risks**



# Impact of REACH



**Registration per chemical: 2M to 10M Euros<sup>1</sup>**

IUCLID: International Uniform Chemical Information Database



**Registrations submitted by 22 Nov. 2010: 19,237<sup>1</sup>**

**Chemicals pre-registered by 1 Dec. 2008: 143,000<sup>2</sup>**



**9M<sup>4</sup> - 54M<sup>3</sup> additional test animals**

*“...promote alternative methods for assessing hazards of substances.....animal testing as a last resort...”* Geert Dancet, ECHA

1. C&EN, Nov 29, 2010, P. 15

2. [http://en.wikipedia.org/wiki/Registration,\\_Evaluation,\\_Authorisation\\_and\\_Restriction\\_of\\_Chemicals](http://en.wikipedia.org/wiki/Registration,_Evaluation,_Authorisation_and_Restriction_of_Chemicals)

3. T. Hartung & C. Rovida: Chemical regulators have overreached. *Opinion in Nature*, vol. 460, 27 Aug '09.

4. ECHA - New study inaccurate on the number of test animals for REACH. Helsinki, 28 August 2009

# Funding



## European Union Seventh Framework Programme



Program: HEALTH-2007-1.3-3

**Promotion, development, validation, acceptance  
and implementation of QSARs for toxicology**

Project Reference Number Health-F5-2008-200787.

**3M Euros (2008 – 2011)**

Awarded to: **OpenTox Consortium**  
**Proposal: “An Open Source Predictive Toxicology Framework”**  
**[www.opentox.org](http://www.opentox.org)**



QSAR: Quantitative Structure-Activity Relationships



# OpenTox Goals



Development of:

- **an interoperable, extensible predictive toxicology framework**

Containing state-of-the-art:

- **QSARs & SARs**
- **cheminformatics, bioinformatics**
- **statistical and data mining tools**
- **computational chemistry & biology algorithms & models**
- ***in vitro* and *in vivo* data resources**
- **ontologies**
- **user interfaces...**”

# The Challenges

- **QSARs for predicting toxicity**

- (Q)SAR algorithms & descriptors
  - consensus predictions
  - validation of models
  - applicability domains
  - transparency



- **Toxicity data**

- standards, quality, licensing, confidentiality

- **Metabolite predictors**

- e.g. *SMARTcyp*, *CypScore*

- **Retrieval of supporting information**

- database mining, text mining

- **Integrated across the internet**





# *in Silico* Methods Considered

**(Q)SAR**

Quantitative Structure-Activity Relationships

**Structural Alerts**

(de)activating fragments or functional groups

**Read Across**

Expert knowledge, compare related chemicals

**Database mining**

for same or similar chemicals

**Text mining**

for reports on same or similar chemicals

**Third-party packages....**



# Reporting: Validation of Predictions

## OECD Guidelines:

### 1. Defined endpoint

well defined 'homogeneous' training data

### 2. Unambiguous prediction algorithm

fully documented prediction models

### 3. Defined applicability domain

documented algorithms for chemical space of model

### 4. Appropriate measures of goodness-of-fit, robustness, & predictivity

scientifically sound:  $r^2$ ,  $CVR^2$ , RMSE, skew, confusion matrix, etc.

### 5. Mechanistic interpretation, if possible

key descriptors, human evaluation, can improve confidence

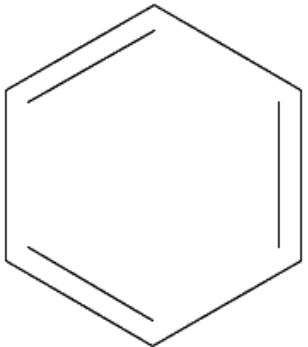
# Prototype Application 1: ToxPredict

## ToxPredict ([www.toxpredict.org](http://www.toxpredict.org))

Browser-based GUI enables lay-user to enter a single structure and get a toxicity report back based on available end-points (currently 18)

The screenshot displays the ToxPredict web application interface. At the top left, there is a 'Help' link. The main heading reads 'Please select the structure(s) for which you would like to apply some OpenTox models.' Below this is a 'Draw' section containing a JME Molecular Editor with a toolbar and a vertical element list (C, N, O, S, F, Cl, Br, I, P, X). To the right of the editor is a 'Search' section with a 'Query\*' input field containing 'benzene', a 'Search mode' section with radio buttons for 'Auto detect' (selected), 'Exact structure', 'Substructure search', and 'Similarity search', and a 'Search' button. At the bottom of the search section, there is a note: 'Enter any identifier (CAS, OpenTox compound or automatically SMILES) alternatively use the JME'. On the right side of the page, there is a navigation menu with sections: 'WELCOME' (Log in), 'PREDICT' (Search structure, Upload structure, View results), 'BROWSE' (Datasets, Models), and 'MY WORKSPACE' (My uploads). At the bottom left, there are social media icons for Twitter, LinkedIn, and a 'More...' link. At the bottom right, it says 'Developed by Ideaconconsult Ltd. 2011'.

# ToxPredict: Results Summary



[Download](#)  
[Browse all](#)  
[Start Read Across](#)

CASRN	71-43-2
EINECS	200-753-7
IUPAC name	benzene
Chemical Name	benzene
SMILES	c1ccccc1
Standard InChI	InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H
Standard InChI key	UHOVQNZJYSORNB-UHFFFAOYSA-N
REACH registration date	30.11.2010

Predictions

**MolecularWeight** [Calculate](#)

MW	78.11
----	-------

**QSAR SRC KOWWIN fingerprints AD** [Calculate](#)

Tanimoto	0.01
AppDomain_Tanimoto	0.00

**Caco-2 Cell Permeability** <http://www.ncbi.nlm.nih.gov/pubmed/16959190> [Calculate](#)

caco2	-4.55
-------	-------

**ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents** [Calculate](#)

At least one positive structural alerts for the micronucleus assay (Class I)	NO
No alerts for the micronucleus assay (Class II)	YES

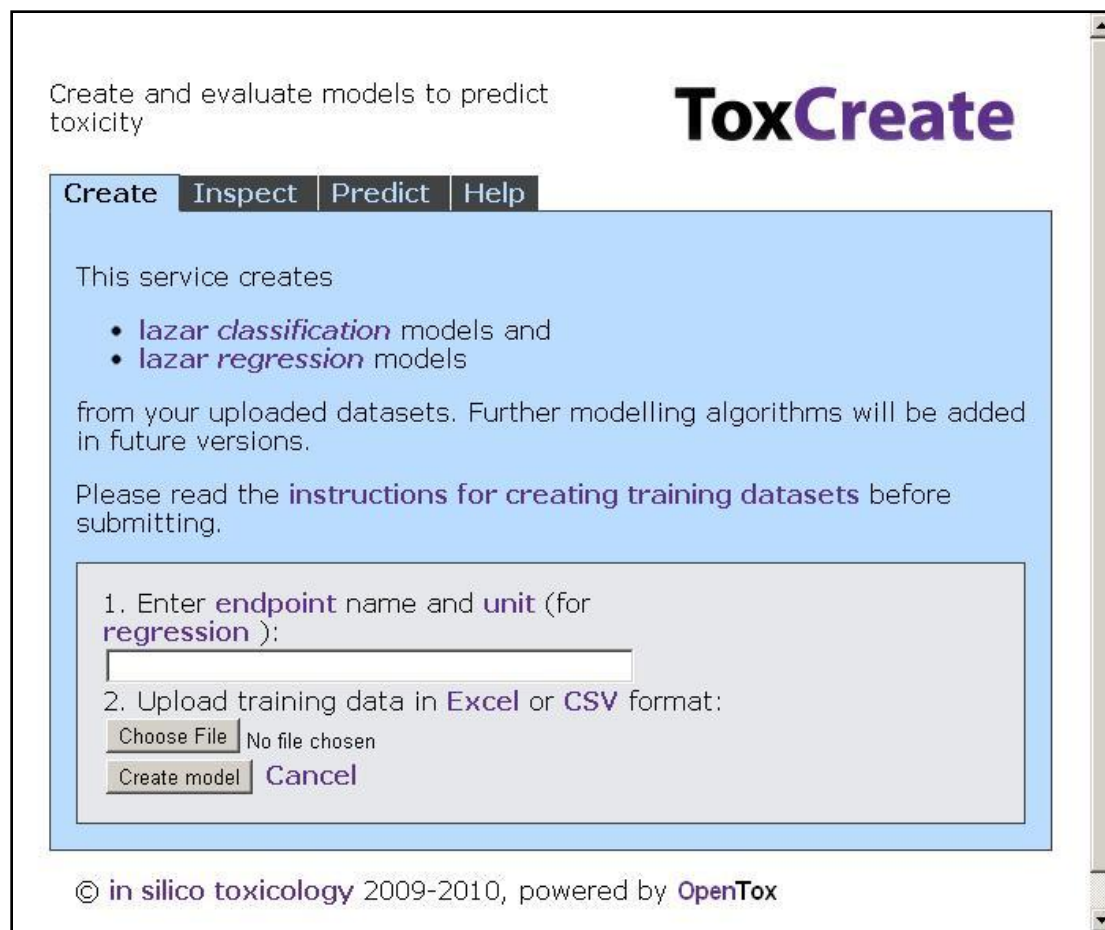
Verbose explanation:

- QSA1.Acyl halides No
- QSA2.Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid No
- QSA3.N-methylol derivatives No
- QSA4.Monohaloalkene No
- QSA5.S or N mustard No
- QSA6.Propiolactones and propiosultones No
- QSA7.Epoxides and aziridines No
- QSA8.Aliphatic halogens No
- QSA9.Alkyl nitrite No
- QSA10.alpha,beta-unsaturated carbonyls No

# Prototype Application 2: ToxCreate

Creates a model from a training set (www.toxcreate.org)

Step 1, upload training set, and create model



Create and evaluate models to predict toxicity

## ToxCreate

Create | Inspect | Predict | Help

This service creates

- *lazar classification* models and
- *lazar regression* models

from your uploaded datasets. Further modelling algorithms will be added in future versions.

Please read the [instructions for creating training datasets](#) before submitting.

1. Enter **endpoint** name and **unit** (for **regression**):
2. Upload training data in **Excel** or **CSV** format:  
 No file chosen

© in silico toxicology 2009-2010, powered by OpenTox

# ToxCreate (2)

## Step 2, View prediction model & validation report

### hamster test2

Status: Completed  
Started: 03/02/2011 - 03:00:18AM  
Training compounds: 85  
Warnings: -  
Algorithm: lazar  
Type: classification  
Descriptors: Fminer backbone refinement  
Training dataset: Excel sheet , RDF/XML (experts, d  
Feature dataset: RDF/XML , YAML (experts, d  
Model: RDF/XML , YAML (experts, m  
Validation:   
Detailed report: [show](#)  
Number of predictions: 76  
Correct predictions: 76.00 %  
Weighted area under ROC: 0.904  
Specificity: 0.771  
Sensitivity: 0.756  
Confusion Matrix:

	Mean
Predicted active	31
Predicted inactive	10

### Crossvalidation report

Created at 03.02.2011 - 03:01

#### Table of Contents

[Mean Results](#)  
[Roc Plot](#)  
[Confusion Matrix](#)  
[Results](#)  
[All Results](#)  
[Predictions](#)

#### Mean Results

This section contains results.

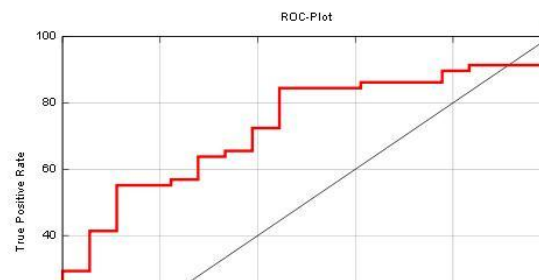
##### Table 1. Mean Results

Algorithm uri	<a href="http://webservices.in-silico.ch/algorithm/lazar">http://webservices.in-silico.ch/algorithm/lazar</a>
Dataset uri	<a href="http://webservices.in-silico.ch/dataset/1702">http://webservices.in-silico.ch/dataset/1702</a>
Num folds	10
Percent correct	69.17 +- 475.31
Weighted area under roc	0.90
Area under roc	true: 0.94, false: 0.89 +- true: 0.03, false: 0.05
F measure	true: 0.78, false: 0.73
True positive rate	0.77
True negative rate	0.77

#### Roc Plot

This section contains the roc plot.

Figure 1. Roc plot



# ToxCreate (3)

## Step 3: enter new structure to predict...

Create and evaluate models to predict toxicity

# ToxCreate

Create | Inspect | **Predict** | Help

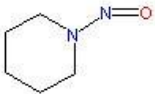
Use this service to obtain predictions from OpenTox models.

Draw a compound

☺ CLR DEL D-R +/- UDO

← — = ≡ ~ △ □ ○ ○ ○ ○

C  
N  
O  
S  
F  
Cl  
Br  
I  
Y  
X



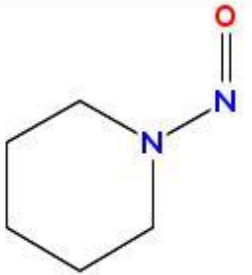
or enter a Name, InChI, Smiles,

Choose one or more prediction models

- hamster test2
- Test BH 6
- hamster carcinogenicity



**New prediction**

O=NN1CCCCC1



hamster test2:  
active  
( Measured activity )

© in silico toxicology 2009-2010, powered by OpenTox



# OpenTox Implementation Outline

- **User Needs**
- **Semantic Web for Predictive Toxicology**
- **Key Components**
- **Web Service Interoperability**
  - Bioclipse, CDK



# Compelling User Needs

## Integrated Testing

*in silico*

*in vitro*

TTC

Read  
Across

Category  
Formation

REACH  
Reporting  
(QPRF, QMRF)

Applicability  
Domain

Validation

Human  
Data

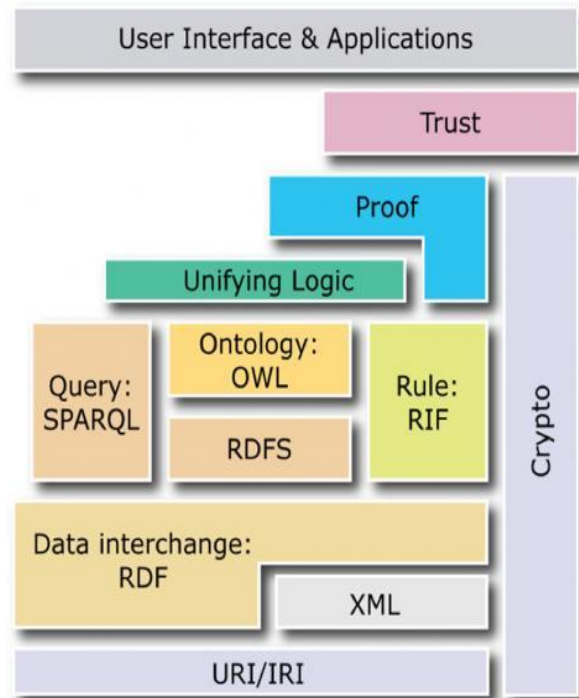
Communicated to OpenTox in 2009 by Grace Patlewicz (Du Pont)

# Semantic Web for Predictive Toxicology

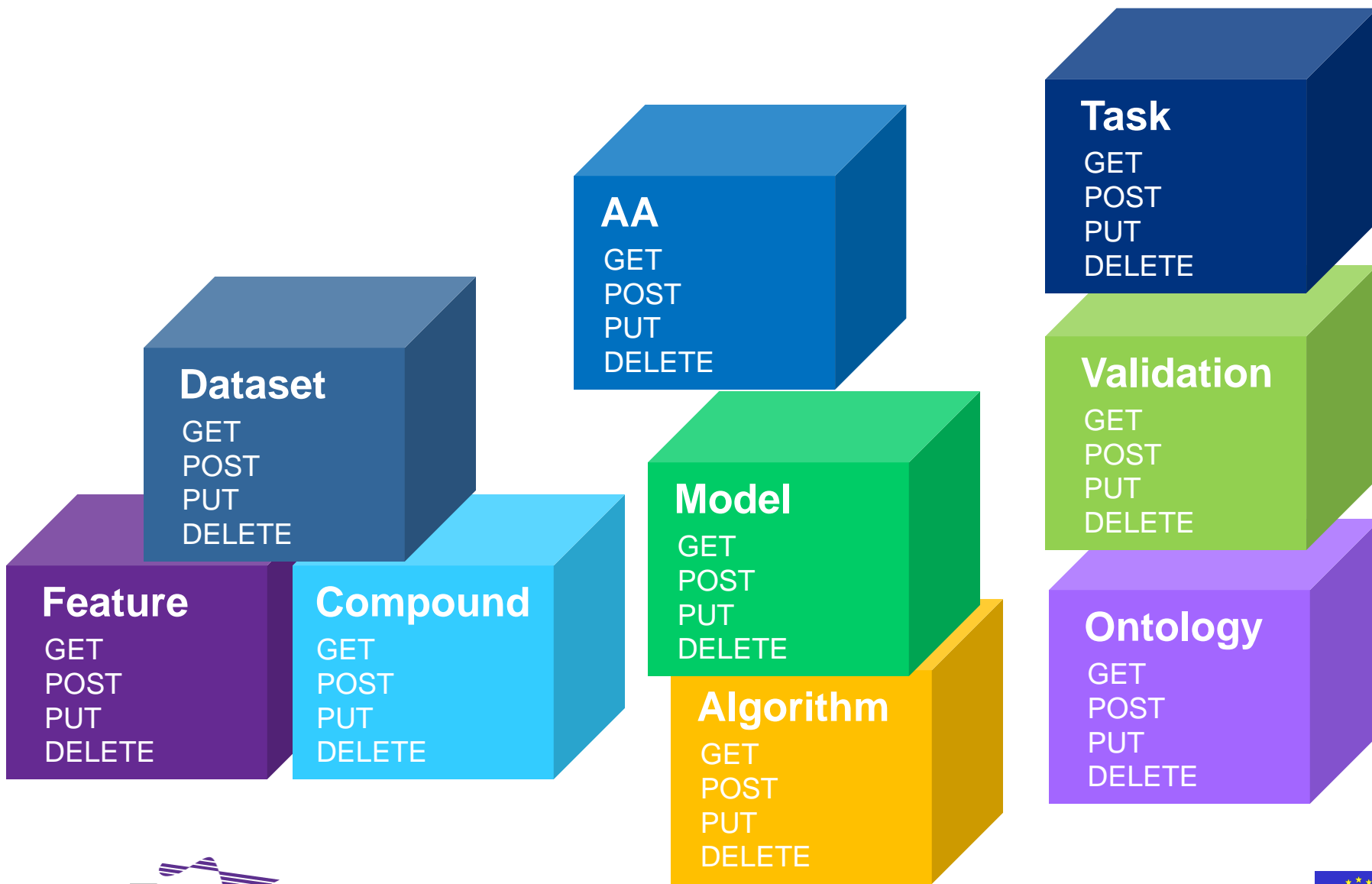
- **URIs:** Uniform Resource Identifier on the web with a name & location
- **RDF:** Resource Description Framework - Graph based Data interchange model to structure data and link resources
- **OWL:** Ontology (Web) Language to add vocabulary for describing properties and relationships



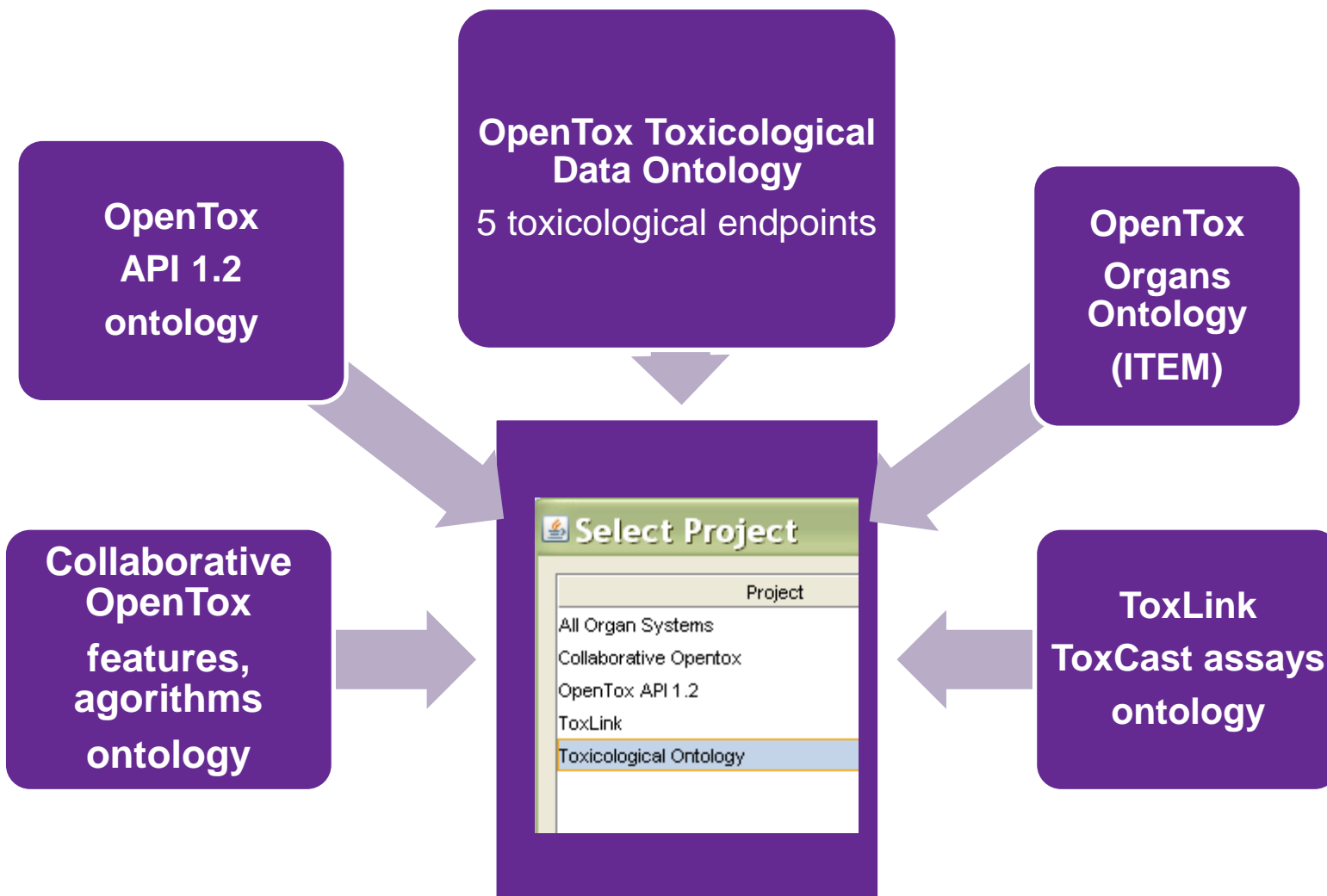
**A www network of Linked Resources for Predictive Toxicology!**



# OpenTox Web Linked Resources/API

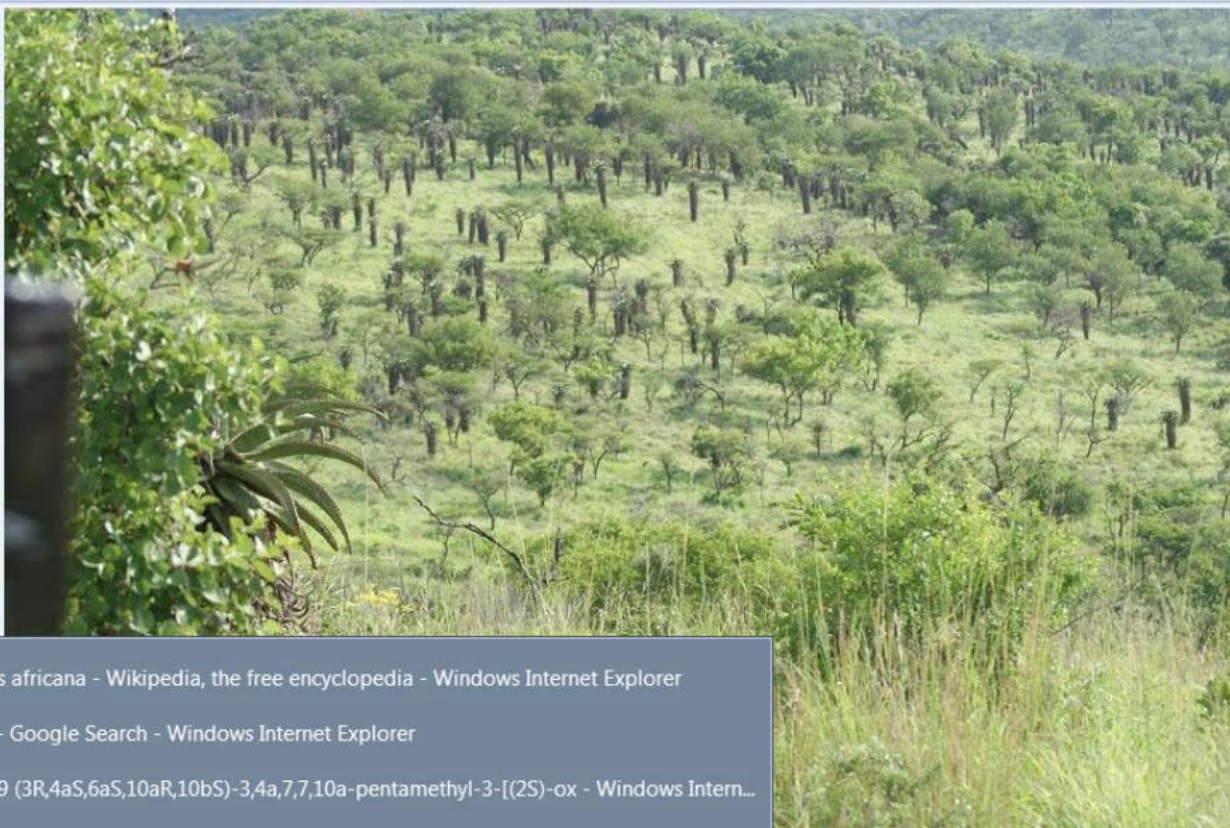


# Collaborative Ontology Development: Collaborative Protege Server



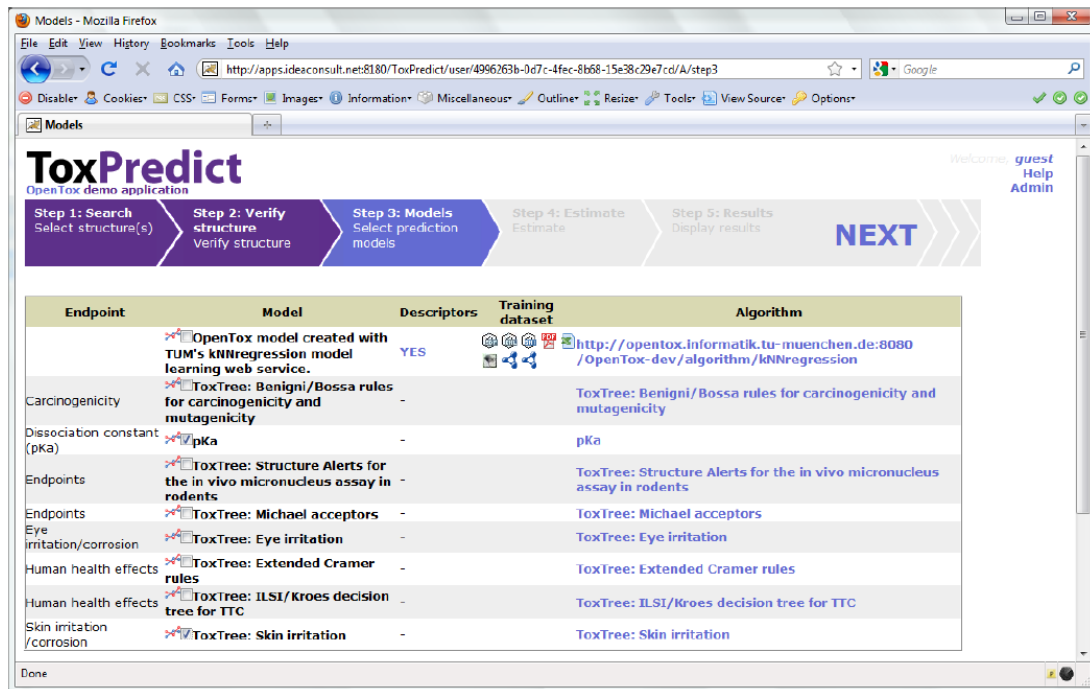
# Web service interoperability

**Bioclipse, CDK**



- W Spirostachys africana - Wikipedia, the free encyclopedia - Windows Internet Explorer
- excoecarin - Google Search - Windows Internet Explorer
- 183900-28-9 (3R,4aS,6aS,10aR,10bS)-3,4a,7,7,10a-pentamethyl-3-[(2S)-ox - Windows Intern...
- Download - Windows Internet Explorer
- GoToMeeting : My Meetings - Windows Internet Explorer

# Recap: What you can do with it ...



Endpoint	Model	Descriptors	Training dataset	Algorithm
	OpenTox model created with TUM's kNNregression model learning web service.	YES	<a href="http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression">http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression</a>	
Carcinogenicity	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	-		ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity
Dissociation constant (pKa)	pKa	-		pKa
Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	-		ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents
Endpoints	ToxTree: Michael acceptors	-		ToxTree: Michael acceptors
Eye irritation/corrosion	ToxTree: Eye irritation	-		ToxTree: Eye irritation
Human health effects	ToxTree: Extended Cramer rules	-		ToxTree: Extended Cramer rules
Human health effects	ToxTree: ILSI/Kroes decision tree for TTC	-		ToxTree: ILSI/Kroes decision tree for TTC
Skin irritation /corrosion	ToxTree: Skin irritation	-		ToxTree: Skin irritation

***Simple building of predictive toxicology applications based on well-established methods and databases***

# Recap: What you can do with it ...



The screenshot shows the ToxPredict web application interface in a Mozilla Firefox browser. The interface includes a navigation bar with three steps: Step 1: Search (Select structure(s)), Step 2: Verify structure (Verify structure), and Step 3: Models (Select prediction models). Below the navigation bar is a table with columns for Endpoint, Model, and Descriptors. The table lists various endpoints and their corresponding models and descriptors.

Endpoint	Model	Descriptors
	OpenTox model created with TUM's kNNregression model learning web service.	YES
Carcinogenicity	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	-
Dissociation constant (pKa)	pKa	-
Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	-
Endpoints	ToxTree: Michael acceptors	-
Eye irritation/corrosion	ToxTree: Eye irritation	-
Human health effects	ToxTree: Extended Cramer rules	-
Human health effects	ToxTree: ILSI/Kroes decision tree for TTC	-
Skin irritation/corrosion	ToxTree: Skin irritation	-

Overlaid on the right side of the screenshot is a map of Europe with five green circular markers placed in various locations across the continent.

**Simple building of applications methods and**

**Distributed applications, integrating wide range of data, models, prediction methods**



# Recap: What you can do with it ...

Endpoint	Model	Descriptors
	OpenTox model created with TUM's kNN regression model learning web service.	YES
Carcinogenicity	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	-
Dissociation constant (pKa)	pKa	-
Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	-
Endpoints	ToxTree: Michael acceptors	-
Eye irritation/corrosion	ToxTree: Eye irritation	-
Human health effects	ToxTree: Extended Cramer rules	-
Human health effects	ToxTree: ILSI/Kroes decision tree for TTC	-
Skin irritation/corrosion	ToxTree: Skin irritation	-

The workflow diagram in Taverna Workbench illustrates a process for model training and testing. It starts with inputting a dataset and an algorithm. The process involves uploading the dataset, training the model, and then applying the trained model to a test set to produce a result. The workflow includes steps like 'calculate\_descriptors', 'get\_features\_of\_trainset', 'learn\_model', 'wait\_for\_learned\_model', 'apply\_model\_to\_testset', and 'wait\_for\_prediction'.

**Simple building of applications methods and**

**Distributed and wide range of methods**

**Integration into workflow systems for computational chemistry & biology**

## **Collaborative development of predictive toxicology applications Journal of Cheminformatics, 2010, 2:7. doi:10.1186/1758-2946-2-7**

Barry Hardy, Nicki Douglas, Christoph Helma, Micha Rautenberg, Nina Jeliaskova, Vedrin Jeliaskov, Ivelina Nikolova, Romualdo Benigni, Olga Tcheremenskaia, Stefan Kramer, Tobias Girschick, Fabian Buchwald, Joerg Wicker, Andreas Karwath, Martin Gutlein, Andreas Maunz, Haralambos Sarimveis, Georgia Melagraki, Antreas Afantitis, Pantelis Sopasakis, David Gallagher, Vladimir Poroikov, Dmitry Filimonov, Alexey Zakharov, Alexey Lagunin, Tatyana Glorizova, Sergey Novikov, Natalia Skvortsova, Dmitry Druzhilovsky, Sunil Chawla, Indira Ghosh, Surajit Ray, Hitesh Patel and Sylvia Escher

Open Access publication: [www.jcheminf.com/content/2/1/7](http://www.jcheminf.com/content/2/1/7)

# OpenTox Achievements to Date:



- ✓ **Integrated services:** prediction models & creation, descriptors, validation
- ✓ **Integrated databases:** ToxCast, ISS, Fraunhofer Inst., Leadscope
- ✓ **Ontologies:** Organs, Toxicological end-points
- ✓ **OpenToxipedia:** terminology related to the study of toxicity
- ✓ **Prototype applications published:** ToxPredict, ToxCreate
- ✓ **API & framework design:** freely available from open source repository
- ✓ **Integration and interoperability:** e.g. OpenTox with Bioclipse & CDK

# OpenTox: The Benefits

- **additional validated toxicity data**

curated databases, new algorithms, search, validation

- **saves time and money**

less animal testing, automated tox reports (OECD, QMRF)

*“ Value statement.....? ”*

# Acknowledgements

**FUNDING:** European Union, 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).

**Douglas Connect, Switzerland (P.I.)**

**In Silico Toxicology, Switzerland**

**Ideaconsult, Bulgaria**

**David Gallagher, UK**

**Seascape Learning & JNU, India**

**Superior Health Institute (ISS), Italy**

**Technical University of Munich, Germany**

**Albert Ludwigs University Freiburg, Germany**

**Fraunhofer Institute for Toxicology & Experimental Medicine, Germany**

**Institute of Biomedical Chem. of the Russian Acad. of Medical Sci., Russia**

**National Technical University of Athens, Greece**

