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.)

OpenTox

SEVENTH FRAMEWORK PROGRAMME

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) 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.

OpenTox

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.)

OpenTox

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

OpenTox

Impact of REACH

Registration per chemical: 2M to 10M Euros[1]

IUCLID: International Uniform Chemical Information Database

Registrations submitted by 22 Nov. 2010: 19,237[1]

Chemicals pre-registered by 1 Dec. 2008: **143,000[2]**

9M[4] - 54M[3] 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
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

OpenTox

OpenTox Goals

Development of:

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

Containing state-of-the-art:

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

OpenTox

QSAR: Quantitative Structure-Activity Relationships
Ontology: formal definitions and relationships between entities in a domain

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**

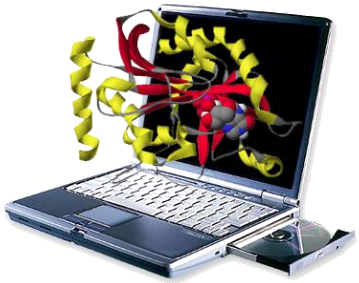
  database mining, text mining

- **Integrated across the internet**

OpenTox

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….	



OpenTox

Reporting: Validation of Predictions

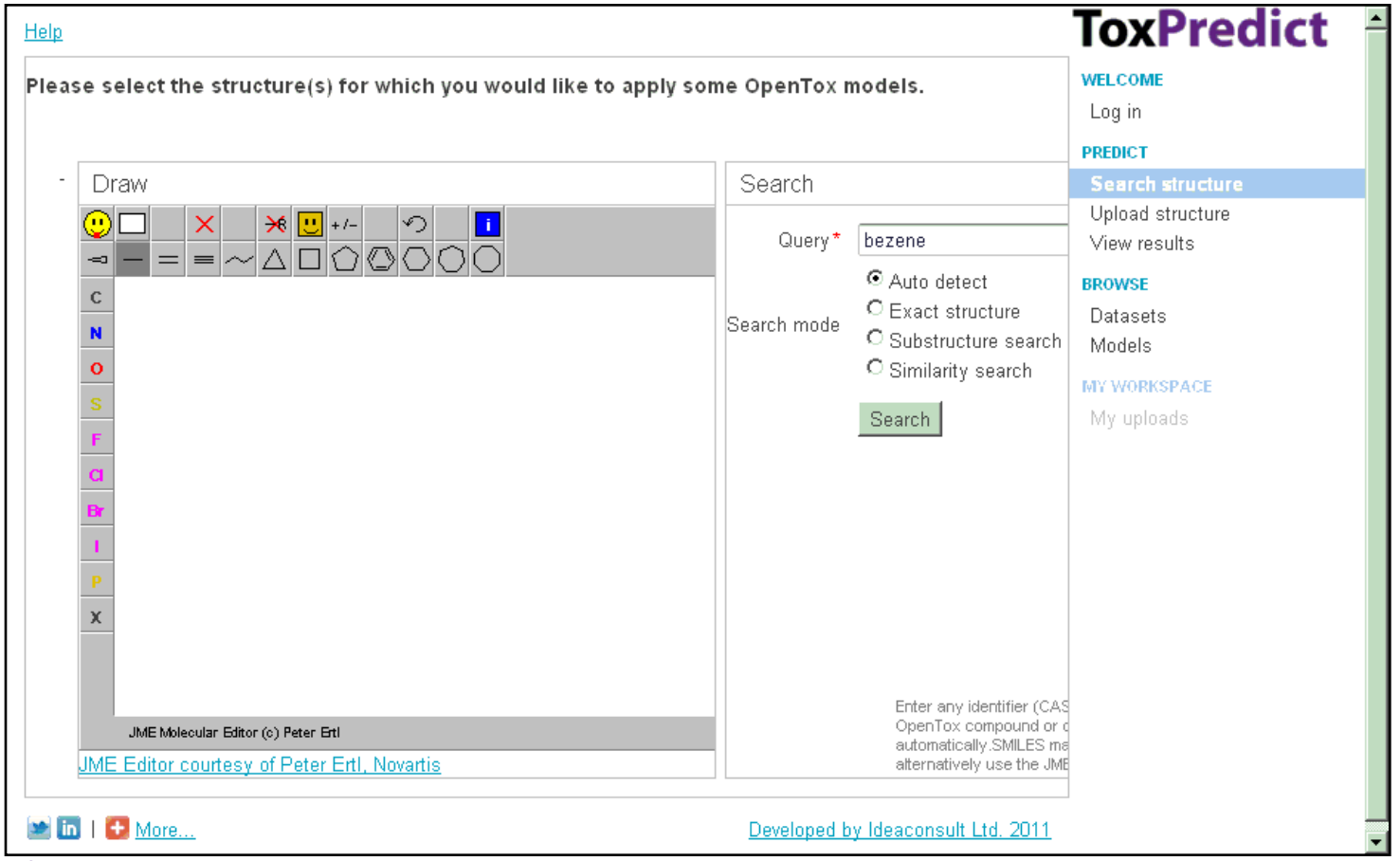
OECD Guidelines:

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

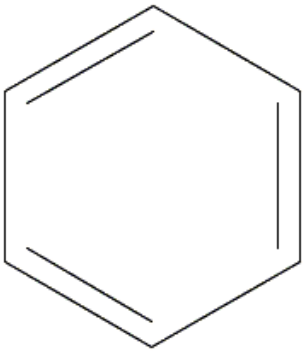
Prototype Application 1: ToxPredict

ToxPredict (www.toxpredict.org)

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



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 | Datasets

Run All

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:

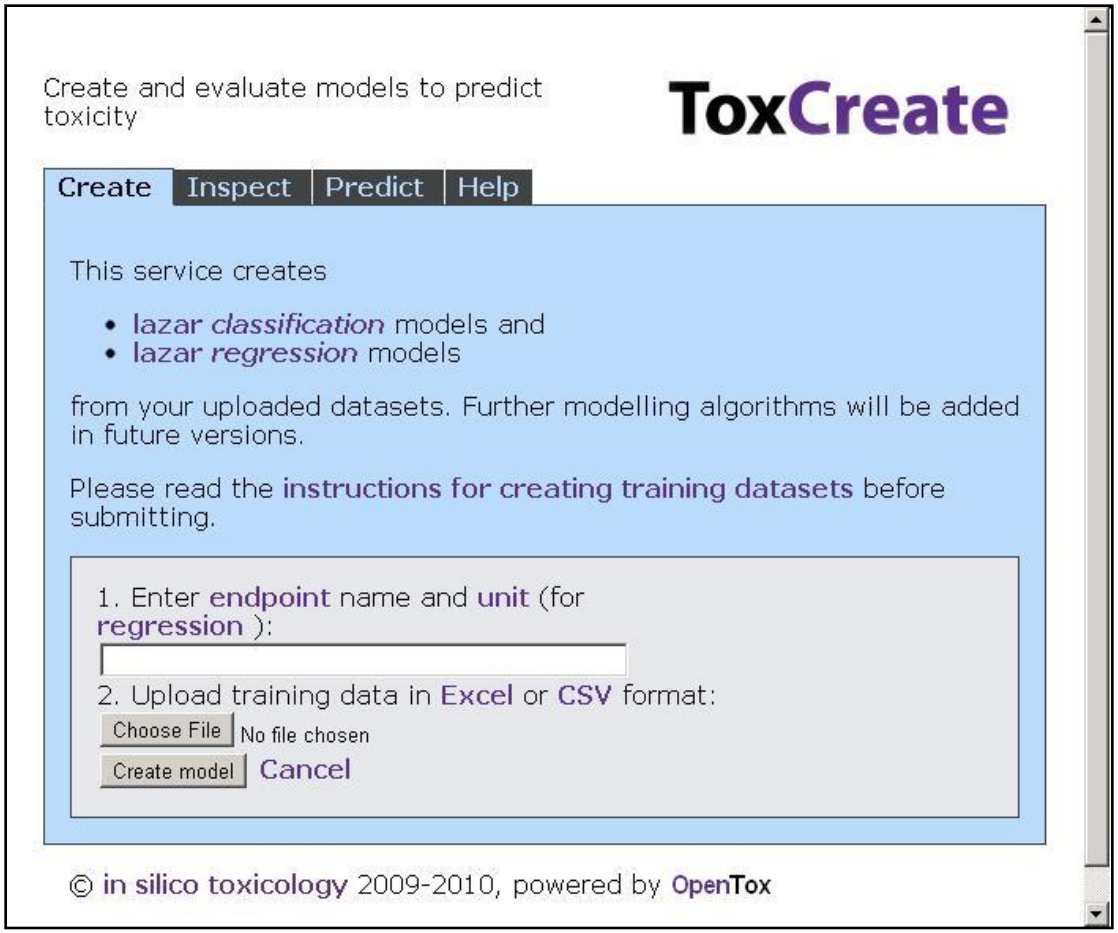
QSA1.Acyl halides No
QSA2.Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid No
QSA3.N-methylol derivatives No
QSA4.Monohaloalkene No
QSA5.S or N mustard No
QSA6.Propiolactones and propiosultones No
QSA7.Epoxides and aziridines No
QSA8.Aliphatic halogens No
QSA9.Alkyl nitrite No
QSA10.α,β-unsaturated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

OpenTox

Prototype Application 2: ToxCreate

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

Step 1, upload training set, and create model



OpenTox

ToxCreate (2)

Step 2, View prediction model & validation report

hampster 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 (expe
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:	

		Mea
		active
Predicted	active	31
	inactive	10

Crossvalidation report

Created at 03.02.2011 - 03:01

Table of Contents

Mean Results

This section contains results.

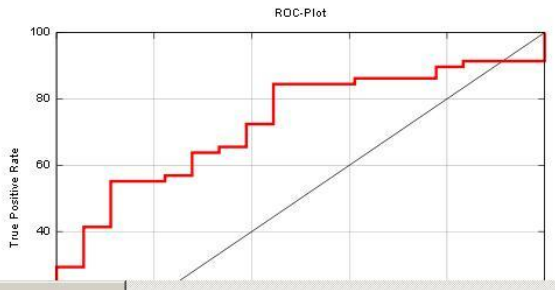
Table 1. Mean Results

Algorithm uri	http://webservices.in-silico.ch/algorithm/lazar
Dataset uri	http://webservices.in-silico.ch/dataset/1702
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.

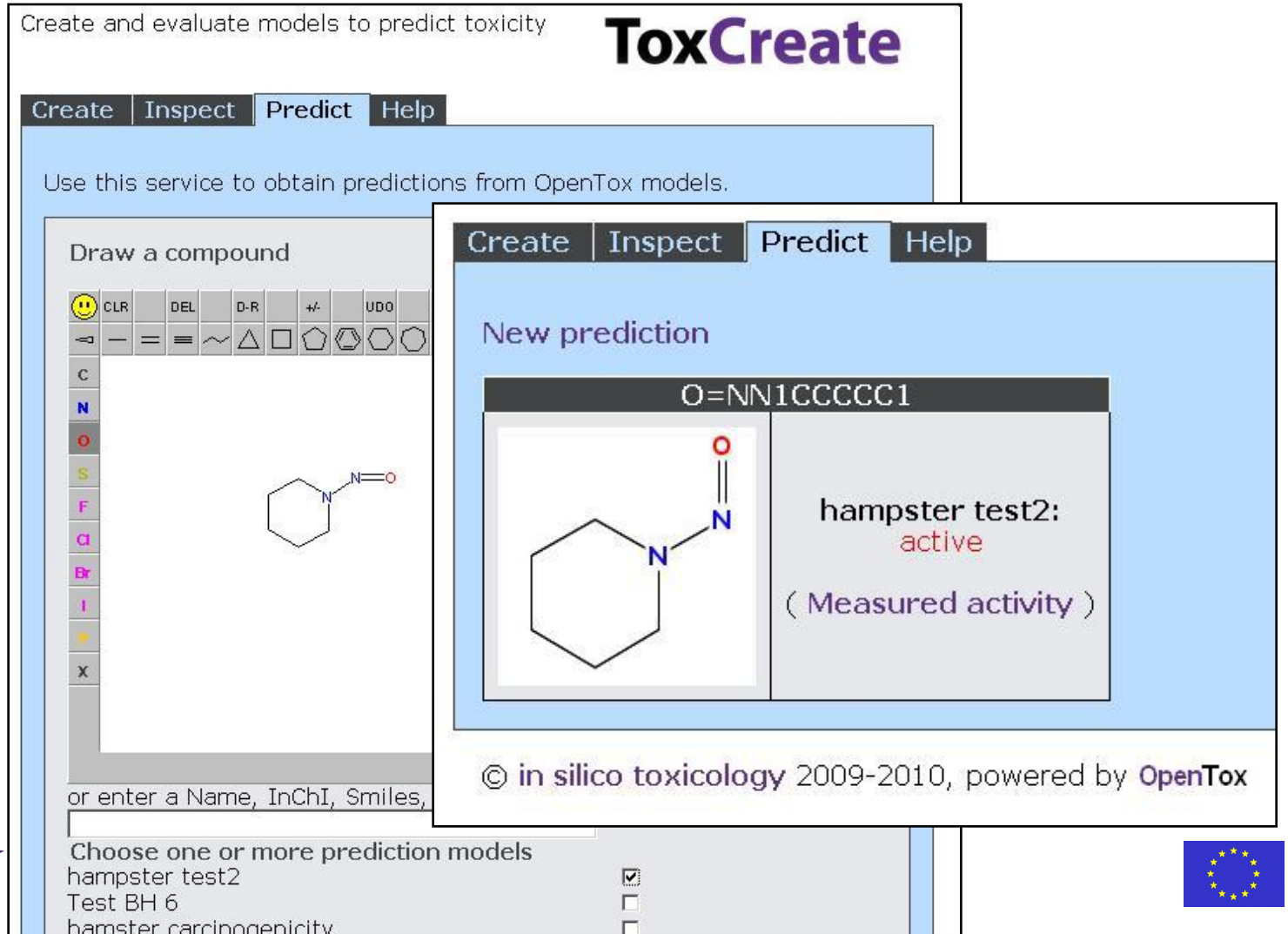
Figure 1. Roc plot



OpenTox

ToxCreate (3)

Step 3: enter new structure to predict…

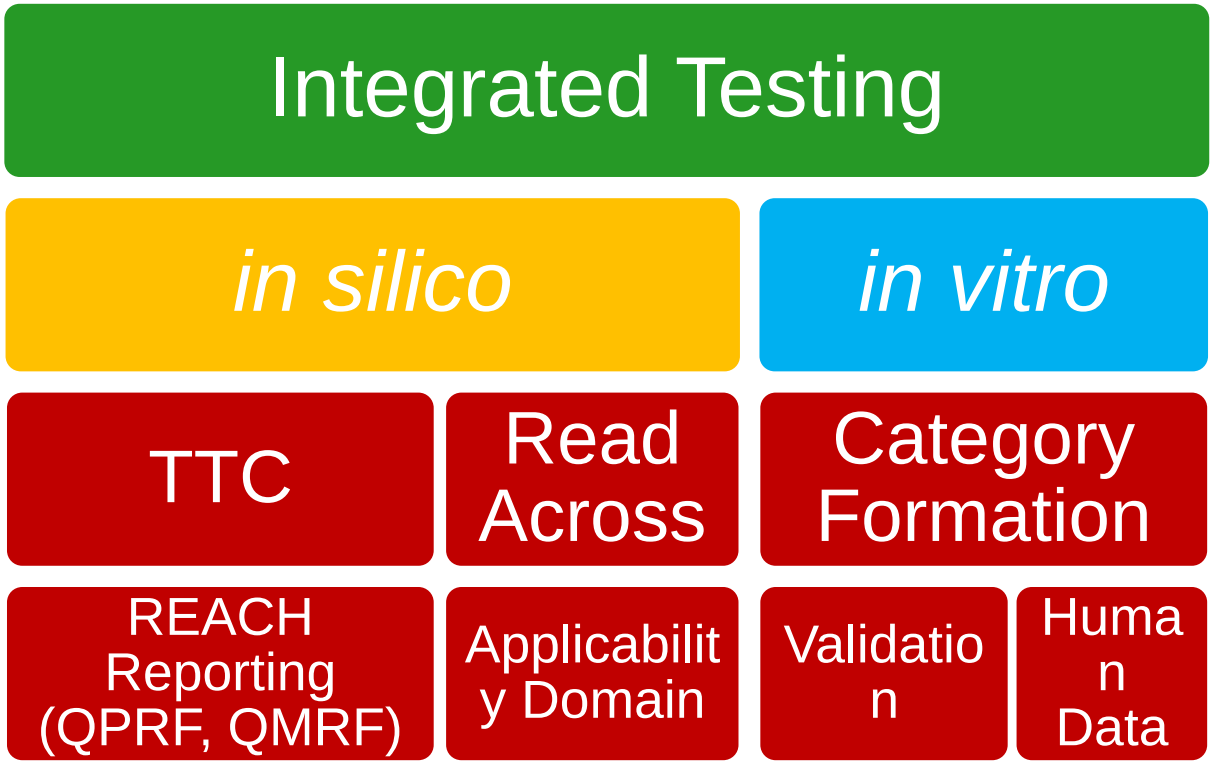


OpenTox

OpenTox Implementation Outline

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

Compelling User Needs

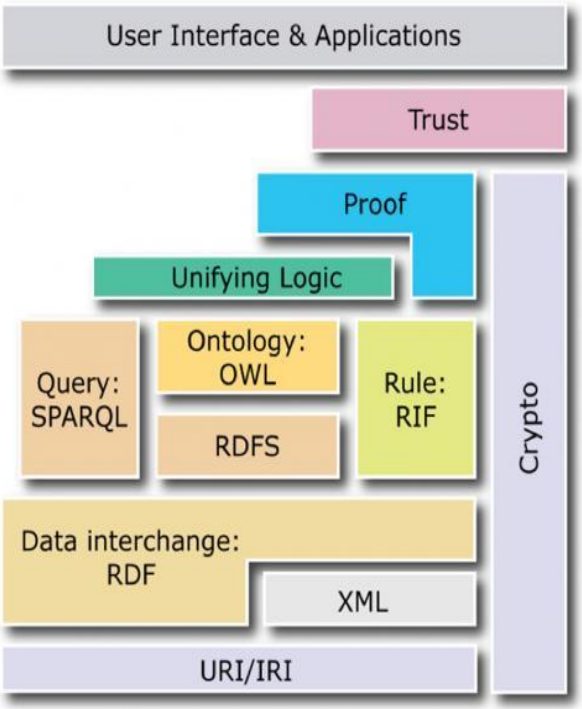


Communicated to OpenTox in 2009 by Grace Patlewicz (Du Pont)

TTC: Threshold of Tox Concern. QPRF: QSAR Prediction Reporting Format.
QMRF: QSAR Model Reporting Format.

Semantic Web for Predictive Toxicology

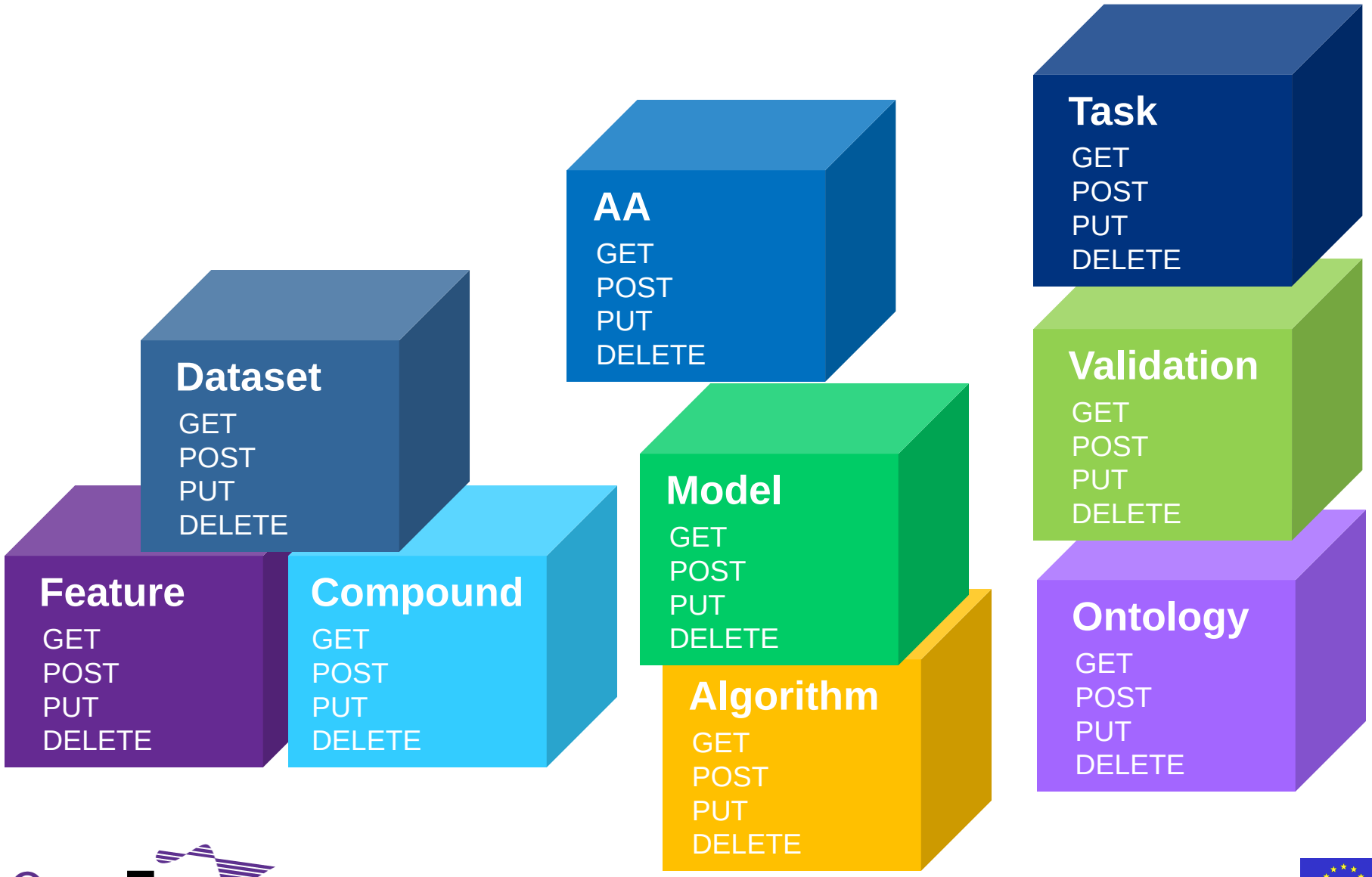
- **URIs**: Uniform Resource Identifier on the web with a name & location
- **RDF:** Resource Description Framework - Graph based Data interchange model to structure data and link resources
- **OWL**: Ontology (Web) Language to add vocabulary for describing properties and relationships



A www network of Linked Resources for Predictive Toxicology!

OpenTox

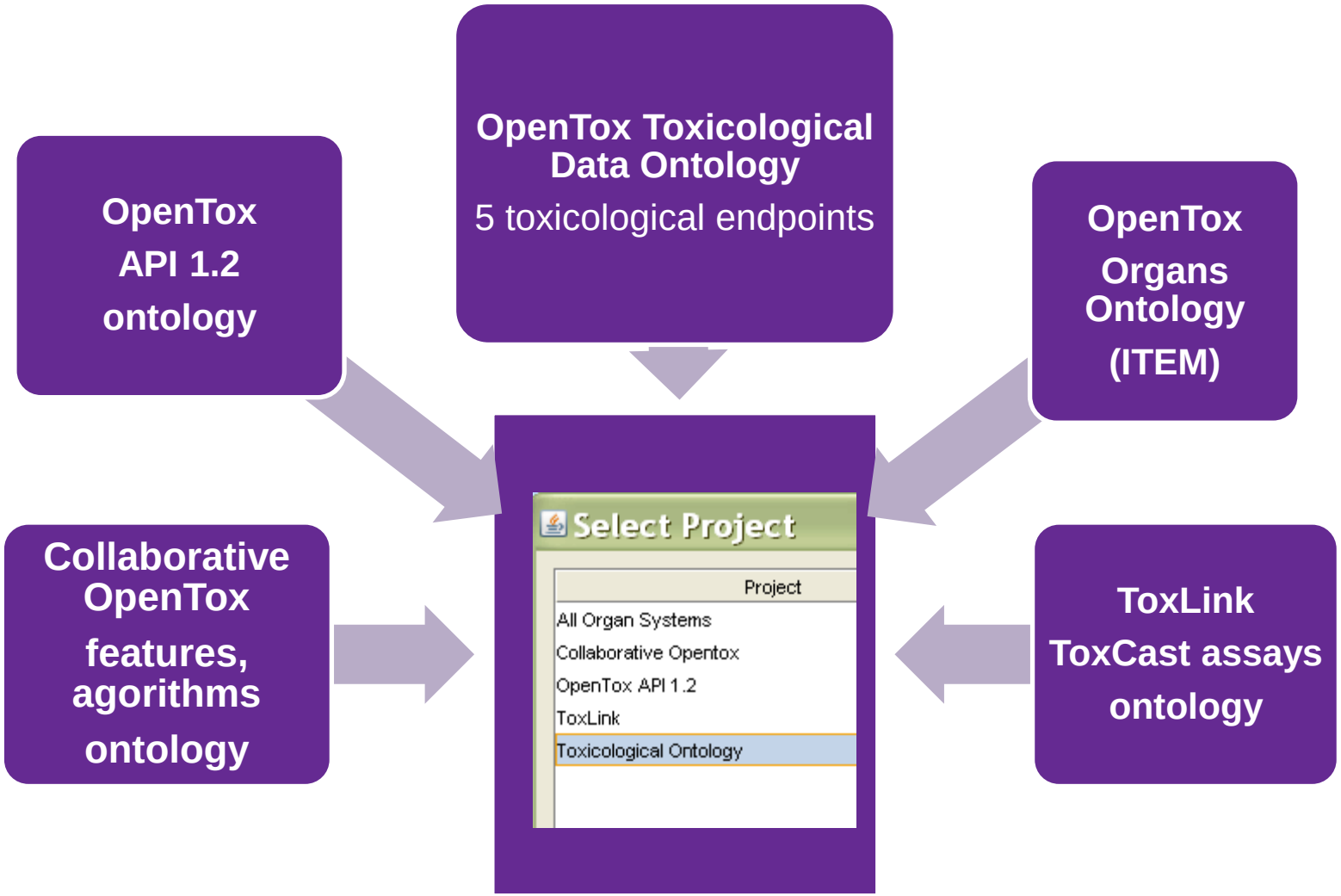
OpenTox Web Linked Resources/API



OpenTox

http://www.Opentox.org/dev/apis/api-1.2

Collaborative Ontology Development: Collaborative Protege Server

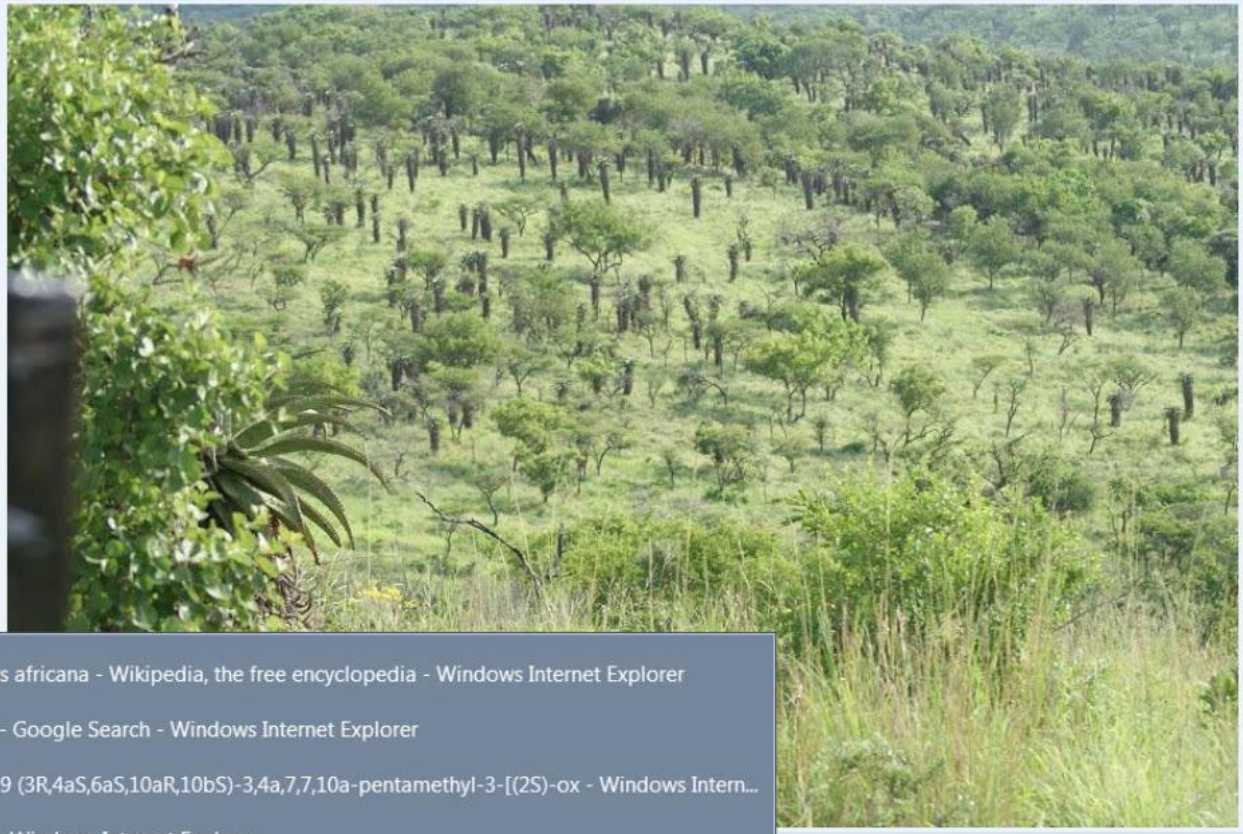


Web service interoperability

Bioclipse, CDK

13 Thanda Bush 101211 1.JPG - Windows Photo Viewer

File Print E-mail Burn Open



Spirostachys africana - Wikipedia, the free encyclopedia - Windows Internet Explorer

excoecarin - Google Search - Windows Internet Explorer

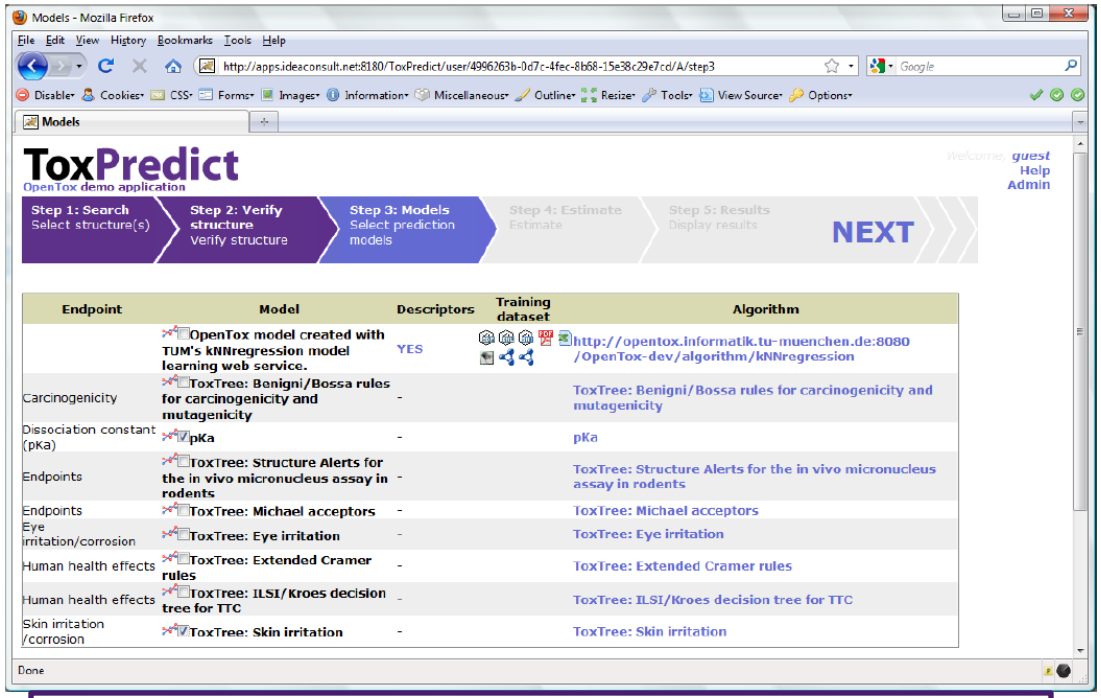
183900-28-9 (3R,4aS,6aS,10aR,10bS)-3,4a,7,7,10a-pentamethyl-3-[(2S)-ox - Windows Intern...

Download - Windows Internet Explorer

GoToMeeting : My Meetings - Windows Internet Explorer

DE 03:42 31.01.2011

Recap: What you can do with it ...



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

OpenTox

Recap: What you can do with it ...

Simple building c

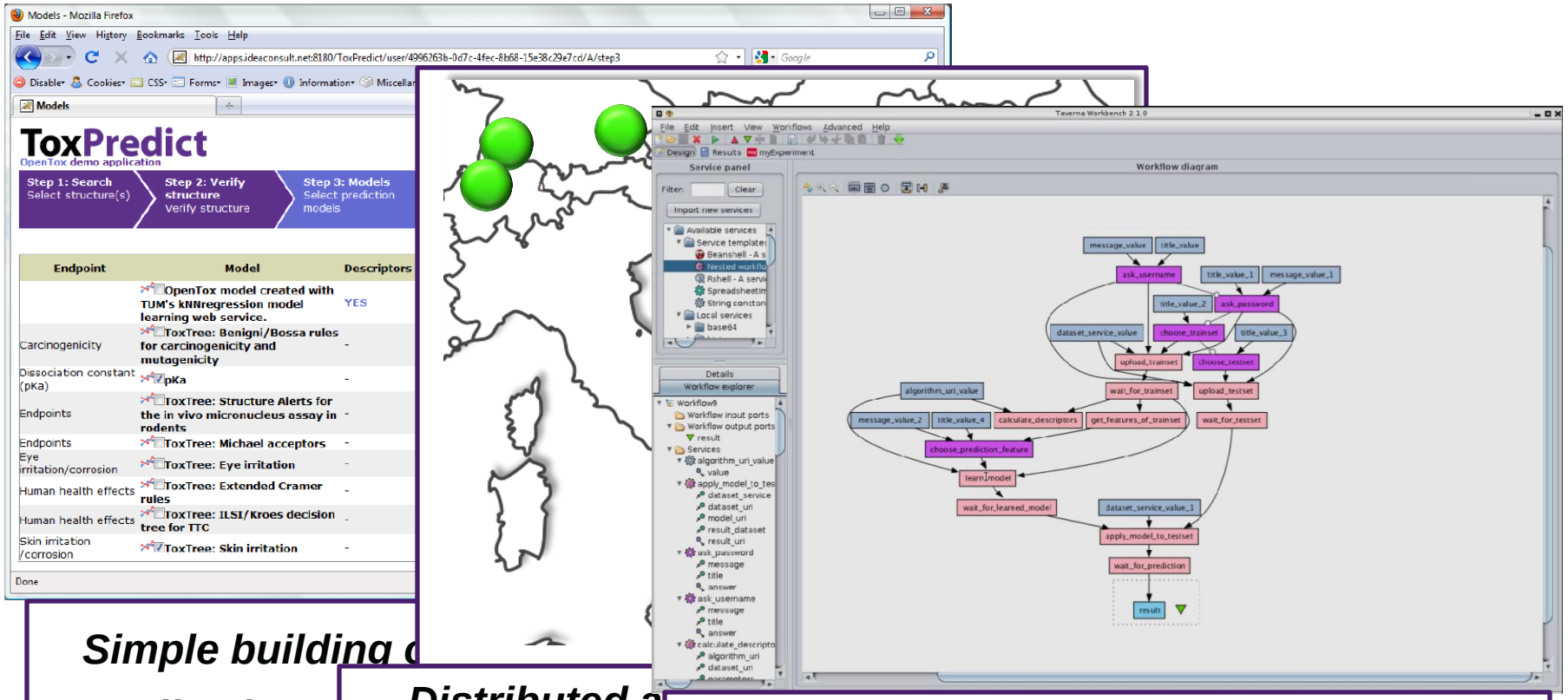
applications

methods and

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

OpenTox

Recap: What you can do with it ...



Simple building

applications

methods and

Distributed a

wide range o

methods

Integration into workflow systems for computational chemistry & biology

OpenTox

Journal of Cheminformatics Publication

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 Jeliazkova, Vedrin Jeliazkov, Ivelina Nikolova, Romualdo Benigni, OlgaTcheremenskaia, Stefan Kramer, Tobias Girschick, Fabian Buchwald, JoergWicker, 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 Gloriozova, 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**

OpenTox

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

API: Application Programming Interface, ISS: Superior Health Institute, Italy
Bioclipse: free, open source workbench for the life sciences.
CDK: open source Java library for Chemoinformatics and Bioinformatics.

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……? ”

OpenTox

API: Application Programming Interface, ISS: Superior Health Institute, Italy
Bioclipse: free, open source workbench for the life sciences.
CDK: open source Java library for Chemoinformatics and Bioinformatics.

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

OpenTox