

OpenTox - An Open Interoperable Predictive Toxicology Framework

Barry Hardy (Douglas Connect)

Barry.Hardy -(at)- DouglasConnect.com 12 April 2011

Current Methods for Computational
Toxicology and Chemogenomics

BIO-IT, Boston, USA

Step 1: Search

Select structure(s)

Step 2: Verify structure

Verify structure

Step 3: Models

Select prediction models


Step 4: Estimate

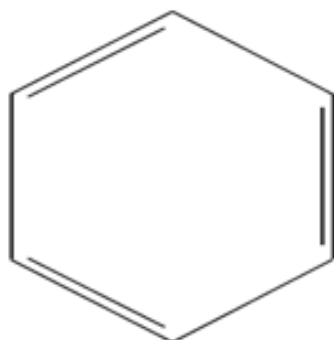
Estimate

Step 5: Results

Display results

This page lists your ToxPredict workflow results for the structure(s) you have selected and the model prediction(s) you have chosen to run. You could also retrieve the ToxPredict report in various other formats, e.g. [SDF](#), [CML](#), [SMI](#), [PDF](#), [CSV](#), [ARFF](#), [RDF/XML](#) or [RDF/N3](#).

Download as 



CAS RN
EINECS
IUPAC name
Synonym

71-43-2
 200-753-7
 benzene
 (6)annulene; benzine; Benzol; Benzolene;
 bicarburet of hydrogen; carbon oil; Coal naphtha;
 cyclohexatriene; mineral naphtha; motor benzol;
 nitration benzene; Phene; Phenyl hydride;
 pyrobenzol.

Synonym
Synonym
Synonym
Quality label

21742.0
 Benzene
 benzene
 OK

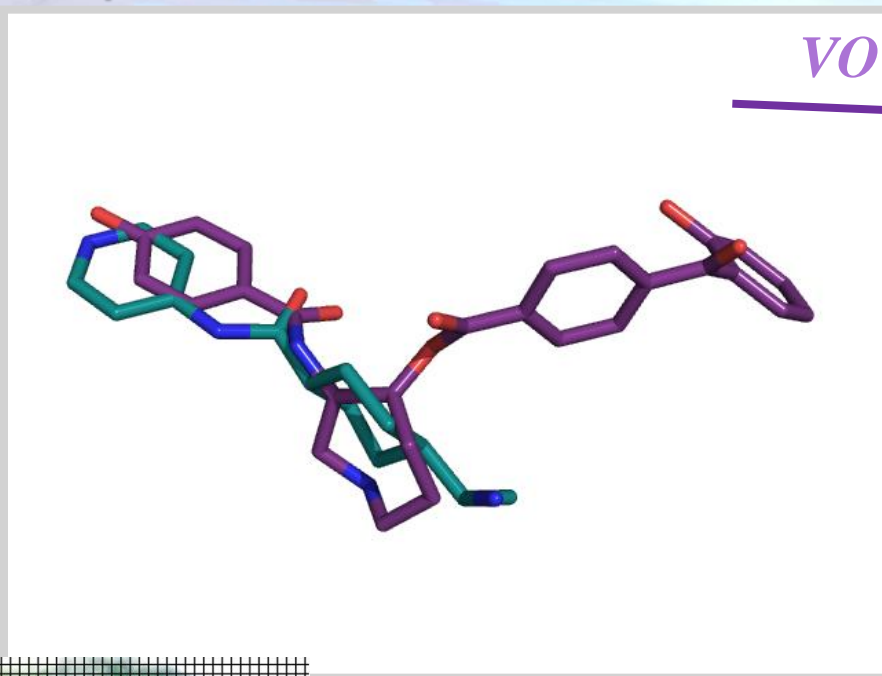
MolecularWeight  **MolecularWeight**

MW

78.1112

Our Use Case

Input Structure



Out - Toxic or Not?

- ☐ LD50
- ☐ Liver Toxicity
- ☐ Secondary Metabolites
- ☐ Interaction with the hERG Channel?
- ☐ Renal Clearance
- ☐ Bioavailability
- ☐ Mutagenicity
- ☐ Carcogenicity
- ☐ Reproductive Toxicology
- ☐ Skin Irritation
- ☐ Aqua Toxicity
- ☐ Combined predictions for arrays of multiple end points



Business Driver

Increasing demands on industry to satisfy safety evaluation and risk assessment required by REACH legislation.

Challenges to *in silico* Applications

- Lack of public standards, ontology
- Toxicity data collected in many different databases using different formats, frequently incompatible with computer programs
- Many databases lack important information for *in silico* modeling (e.g. chemical structures)
- Hard to integrate confidential in-house data with public data for model building and validation
- Models have been published in a variety of different formats (ranging from simple regression equations to complete computer programs)
- Need for New Business Cases
- There is no straightforward integration of predictions from various programs
- No commonly accepted framework for validation of *in silico* predictions, many tools provide limited support for reliable validation procedures
- Application, interpretation, and development of *in silico* models is still difficult for most toxicological experts
- It requires a considerable amount of statistical, cheminformatics and computer science expertise - procedures are labor intensive and prone to human errors

Compelling Needs of Users

Multidisciplinary R&D

Good Support of Flexible Applications

Transparency -
Not Black Box!

Mechanistic
rationale

QSAR &
Expert
Systems

Workflows

Automated
Integration

Applicability
Domain

Categories

Systems
Biology

Compelling Needs of Users

Integrated Testing

in silico

in vitro

TTC

Read
Across

Category
Formation

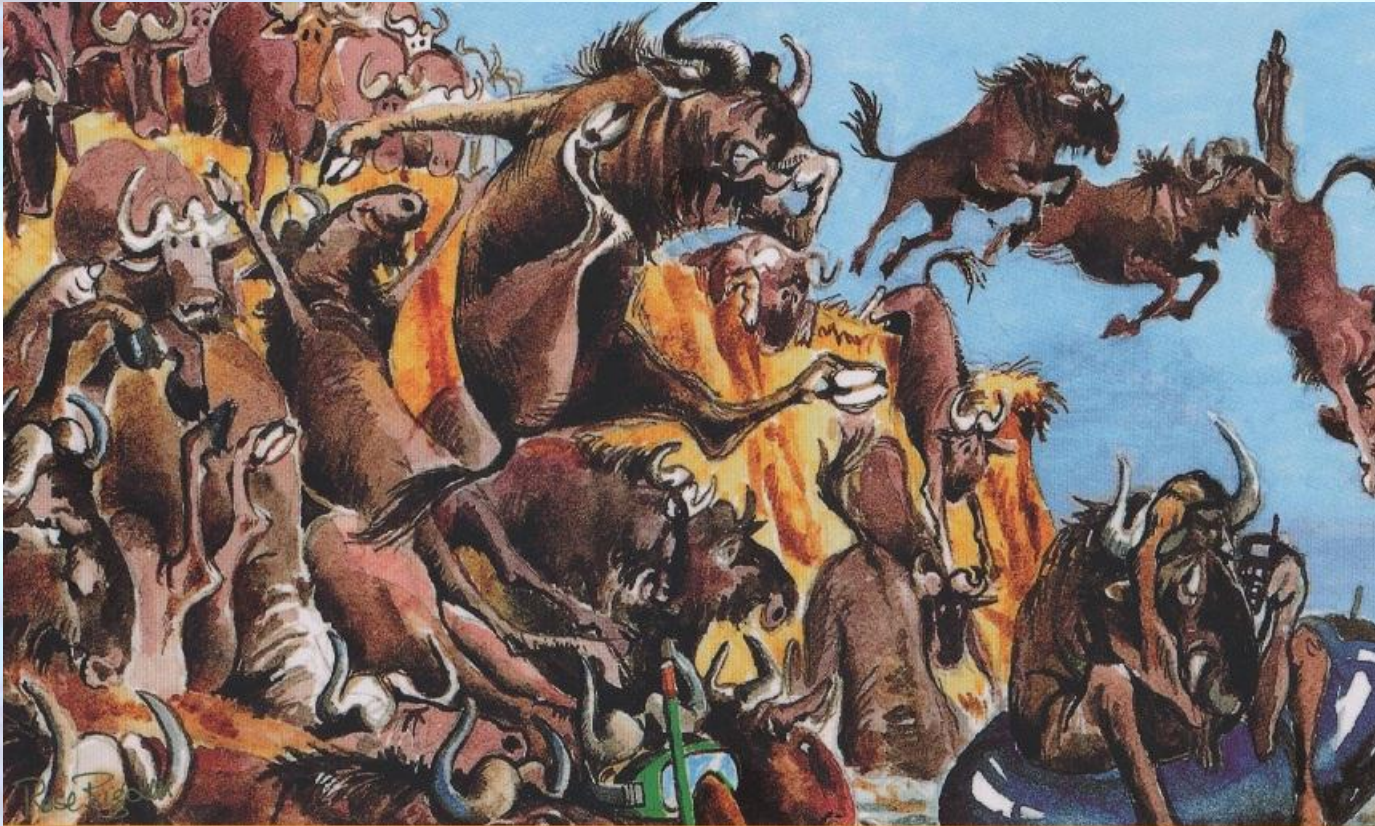
REACH Reporting
(QPRF, QMRF)

Applicability
Domain

Validation

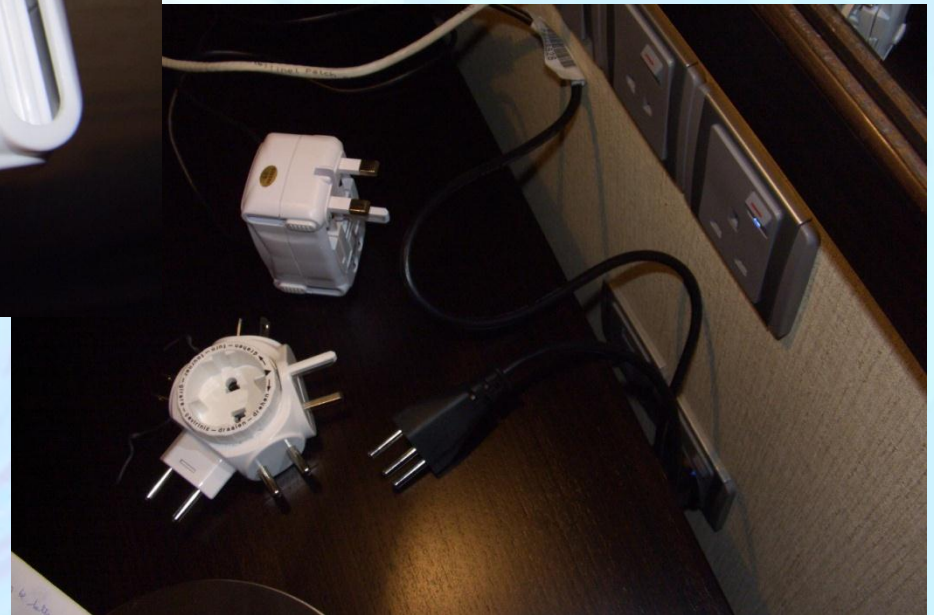
Human
Data

Migration



I'll get back to you, I'm in the middle of this migration thing!

Ontology and Data - Interoperability



Interacting Components create Solutions



Adaptor Solution in Jeddah, 2008

OpenTox Components

Compounds: Structures, names, ...

Features: Chemical and biological (toxicological) properties, substructures, ...

Datasets: Relationships between compounds and features

Algorithms: Instructions for solving problems

Models: Algorithms applied to data yield models which can be used for predictions

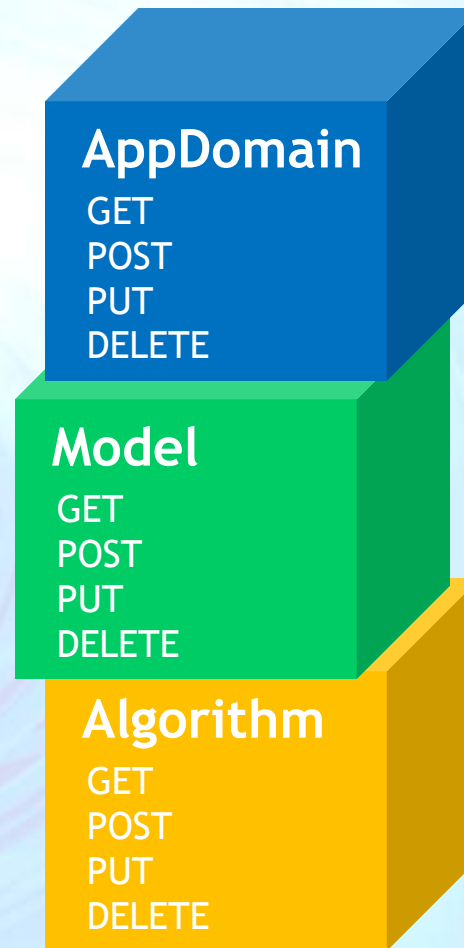
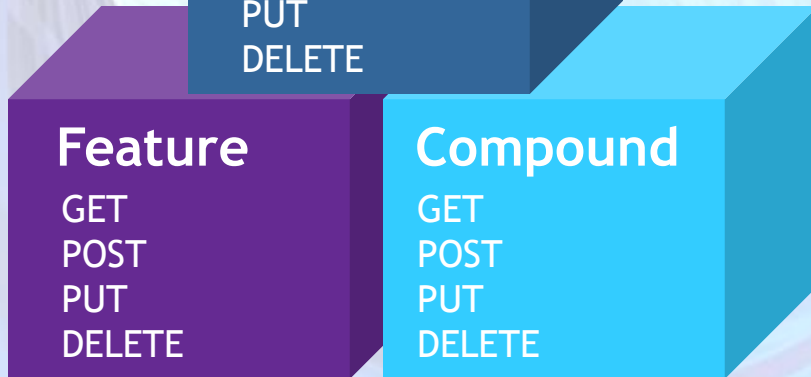
Validation: Methods for estimating the accuracy of model predictions

Reports: Report predictions and models e.g. to regulatory authorities

Tasks: Handle long running calculations

Authentication and Authorisation: Protect confidential data

Overview of Application Programming Interfaces



Representational State Transfer (REST)

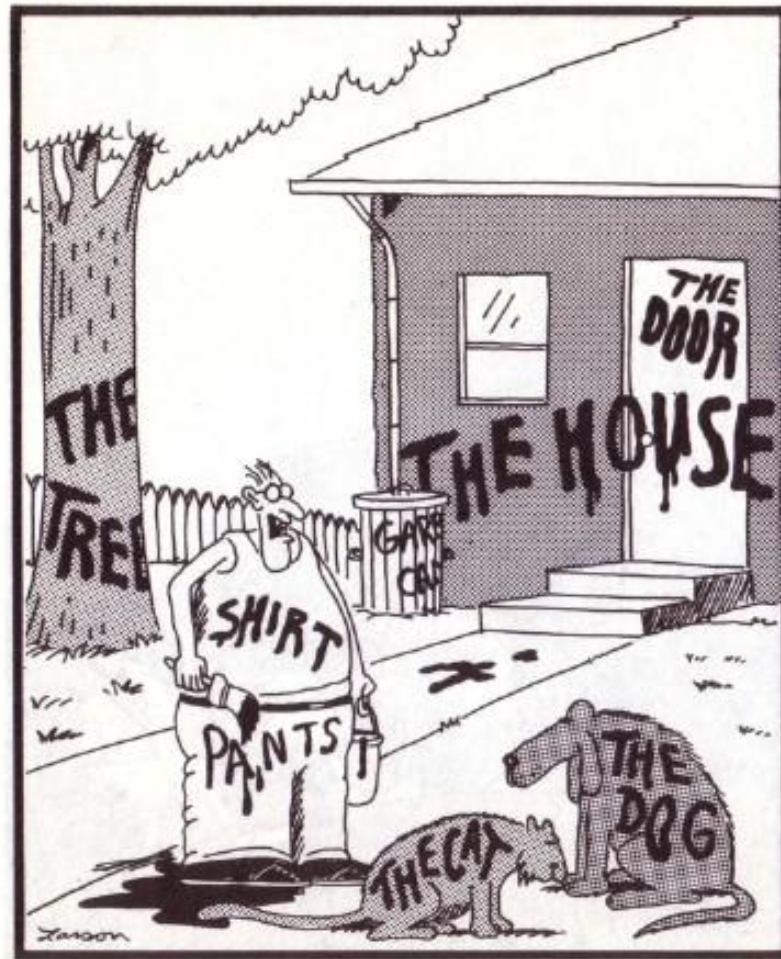
What?

- Architectural style for distributed information systems on the Web
- Simple interfaces, data transfer via **hypertext transfer protocol (HTTP)**, stateless client/server protocol
 - GET, POST, PUT, DELETE
- Each **resource** is **addressed** by its own **web address**

Why?

- **Lightweight** approach to **web services**
- **Simplifies/enables** development of **distributed and local systems**
- Language independent

Semantic Reflections



"Now! ... That should clear up
a few things around here!"

Interoperability & Vocabulary

Dogs

- Collie
- Labrador



Cats

- Siamese
- Persian

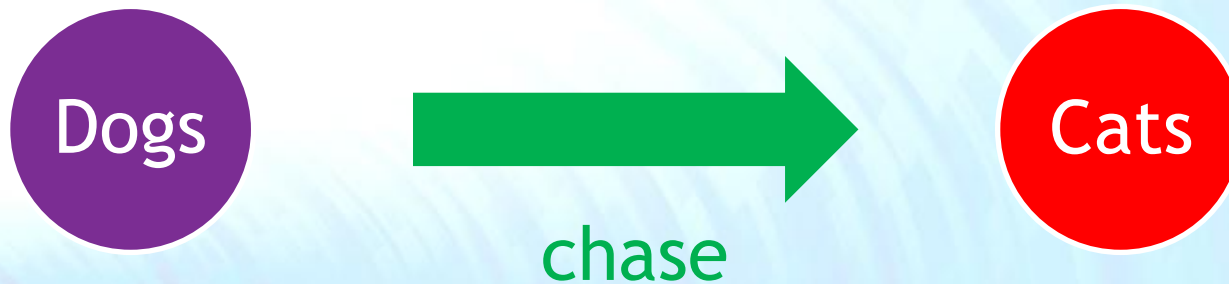


Birds

- Sparrow
- Owl



Interoperability & Vocabulary



Interoperability & Ontology

Org A

Collie

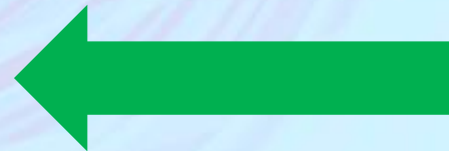


Siamese



Org B

Collie



Siamese

OpenTox committed to creating a Semantic Web for Predictive Toxicology (*with its API 1.1 development in 2009*)

Linked Data is a term used to describe the exposing, sharing, and connecting of data on the Semantic Web using:

URIs a generic means to identify entities in the world

HTTP a simple yet universal mechanism for retrieving resources

RDF a generic graph-based data model with which to structure and link data

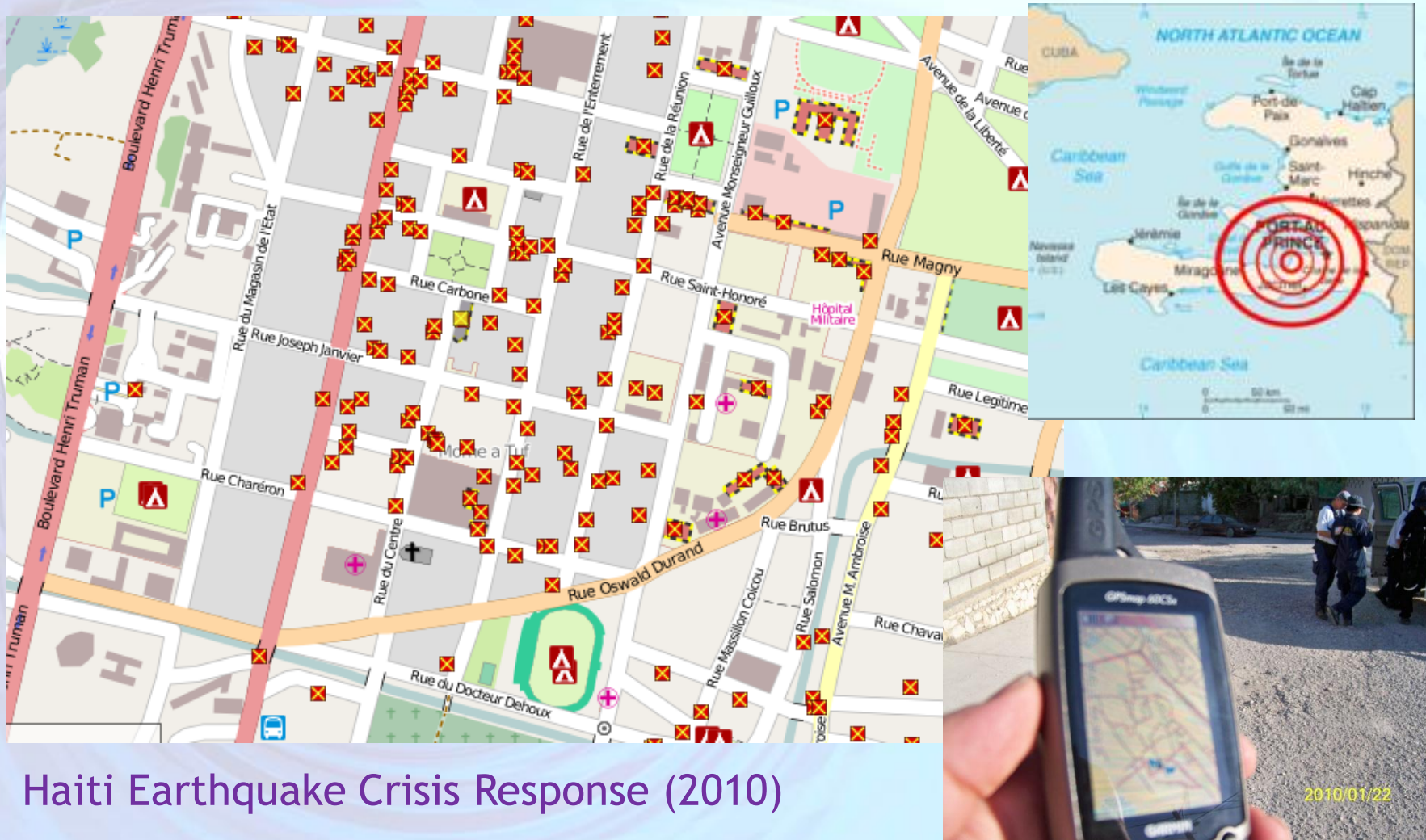
Linked Data needs:

1. Provision of a **URI** that describes a Data Resource
2. Use of **HTTP** to retrieve useful data from the **URI**
3. A Data Format described with standardised semantics (so relationships are enabled) e.g. **RDF**
4. Data should provide links to other Data (through **URIs**)



DBpedia = Linked Data approach applied to Wikipedia

Solution created by Linked Open Data, Web Applications and Crowdsourcing



Haiti Earthquake Crisis Response (2010)

wiki.openstreetmap.org

	OECD Principle	OpenTox addresses Validation Principles by...
1	Defined Endpoint	providing a unified source of well defined and documented toxicity data with a common vocabulary
2	Unambiguous Algorithm	providing transparent access to well documented models and algorithms as well as to the source code
3	Defined Applicability Domain	integrating tools for the determination of applicability domains during the validation of prediction models
4	Goodness-of-fit, robustness and predictivity	providing scientifically sound validation routines for the determination of errors and confidences
5	Mechanistic interpretation (if possible)	integrating tools for the inference, correlation or prediction of toxicological mechanisms and the recording of opinions and analysis in reports

OpenTox is an Integrating Framework

Framework

- Toxicity Data (Linked)
- *in silico* models
- Validation & Reporting
- Interpretation aids

Diverse Access

- Toxicologist, Biologist, Chemists
- Computational Scientists
- Interfaces for new algorithm development & integration

Interoperability

- Promote Standards
- Core Open Source Components
- Support Ontologies & Integration of Multiple Resources

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 Jeliaskova, Vedrin Jeliaskov, 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 Glorizova, Sergey Novikov, Natalia Skvortsova, Dmitry Druzhilovsky, Sunil Chawla, Indira Ghosh, Surajit Ray, Hitesh Patel and Sylvia Escher

Open Access publication available at
www.jcheminf.com/content/2/1/7

Step 1: Search

Select structure(s)

Step 2: Verify structure

Verify structure

Step 3: Models

Select prediction models


Step 4: Estimate

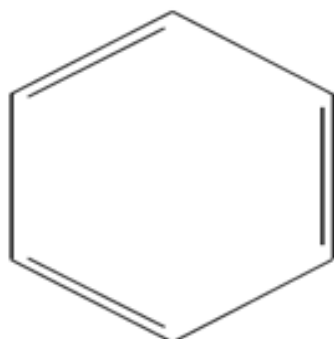
Estimate

Step 5: Results

Display results

This page lists your ToxPredict workflow results for the structure(s) you have selected and the model prediction(s) you have chosen to run. You could also retrieve the ToxPredict report in various other formats, e.g. [SDF](#), [CML](#), [SMI](#), [PDF](#), [CSV](#), [ARFF](#), [RDF/XML](#) or [RDF/N3](#).

Download as 



CAS RN
EINECS
IUPAC name
Synonym

71-43-2
 200-753-7
 benzene
 (6)annulene; benzine; Benzol; Benzolene;
 bicarburet of hydrogen; carbon oil; Coal naphtha;
 cyclohexatriene; mineral naphtha; motor benzol;
 nitration benzene; Phene; Phenyl hydride;
 pyrobenzol.

Synonym
Synonym
Synonym
Quality label

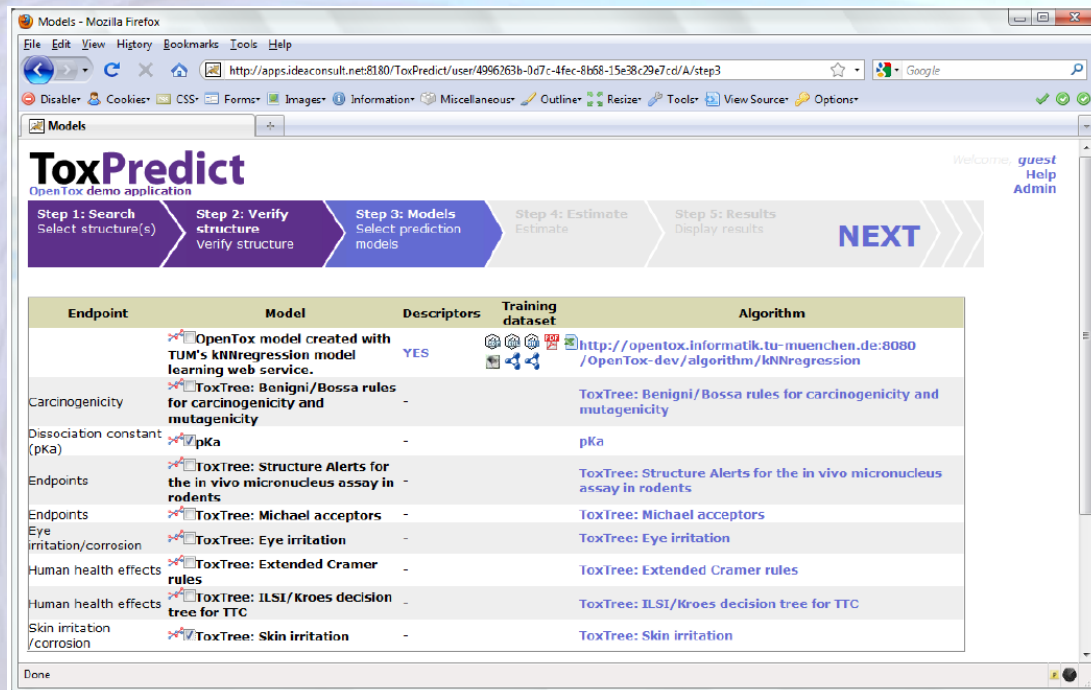
21742.0
 Benzene
 benzene
 OK


MolecularWeight  **MolecularWeight**

MW

78.1112

What you can do with it ...



Endpoint	Model	Descriptors	Training dataset	Algorithm
	OpenTox model created with TUM's kNNregression model learning web service.	YES	 http://opentox.informatik.tu-muenchen.de:8080/OpenTox-dev/algorithm/kNNregression	
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

What you can do with it ...



The screenshot shows the ToxPredict web application in a Mozilla Firefox browser. The URL is <http://apps.ideaconsult.net:8180/ToxPredict/user/4996263b-0d7c-4fec-8b58-15e38c29e7cd/A/step3>. 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 this is a table with columns for Endpoint, Model, 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	-

Below the table is a "Done" button. To the right of the browser window is a map of Europe with five green circular markers placed in various locations: two in Northern Europe (UK/Ireland area), one in Central Europe (Germany/Poland area), one in Eastern Europe (Russia/Ukraine area), and one in Southern Europe (Spain/Portugal area).

Simple building of applications methods and

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

What you can do with it ...

The screenshot displays the ToxPredict application interface. On the left, a table lists various endpoints and their corresponding models. On the right, a workflow diagram illustrates the process of building and using a model.

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	-

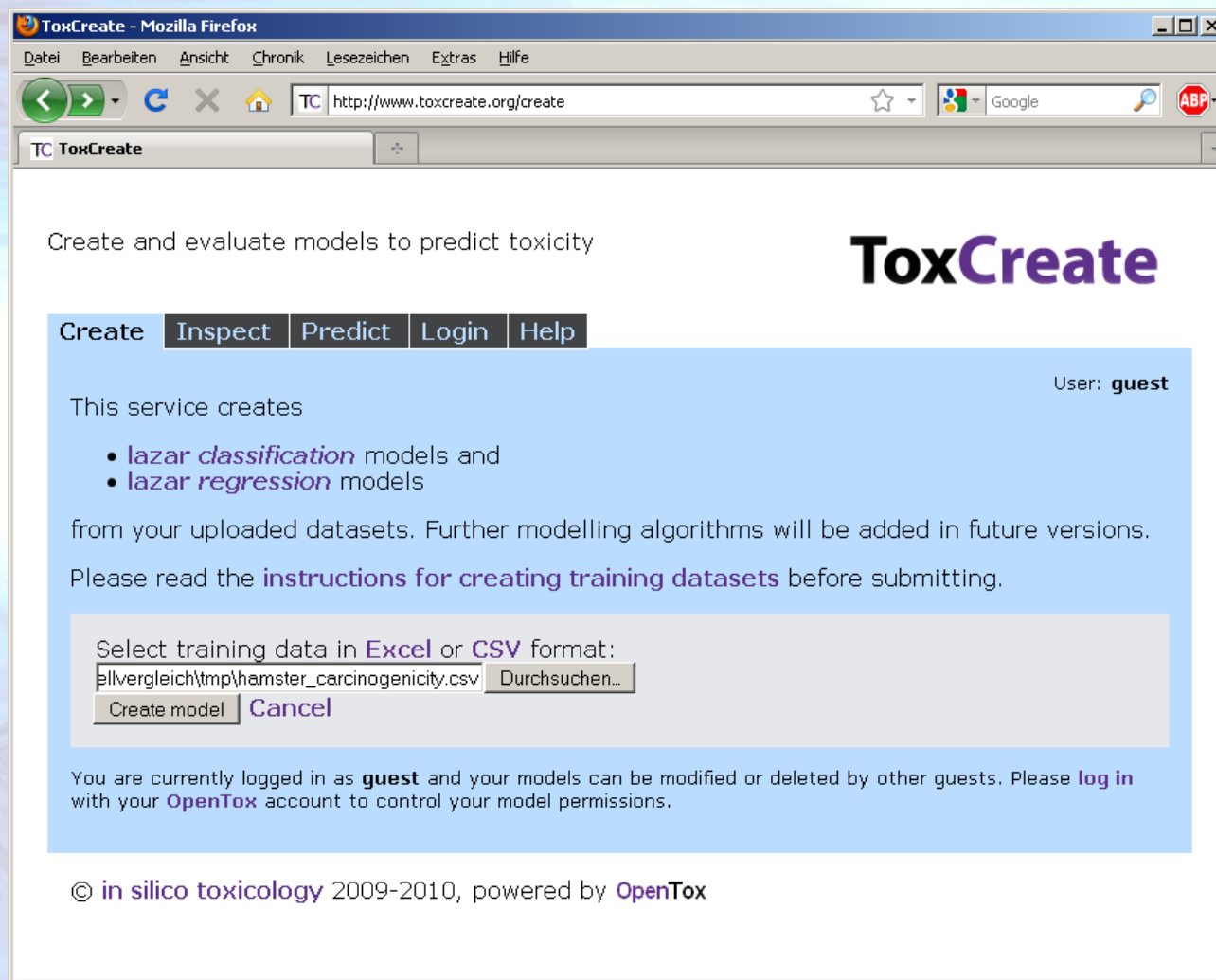
The workflow diagram on the right shows a sequence of steps: **ask_username**, **ask_password**, **choose_trainer**, **choose_testset**, **upload_trainer**, **upload_testset**, **wait_for_trainer**, **wait_for_testset**, **calculate_descriptors**, **get_features_of_trainer**, **wait_for_trainer**, **choose_prediction_feature**, **learn_model**, **wait_for_learned_model**, **dataset_service_value_1**, **apply_model_to_testset**, **wait_for_prediction**, and **result**.

Simple building of applications methods and

Distributed of wide range of methods

Integration into workflow systems for computational biology

ToxCreate



The screenshot shows a Mozilla Firefox browser window with the address bar displaying <http://www.toxcreate.org/create>. The page title is "ToxCreate". The main heading is "Create and evaluate models to predict toxicity". Below this, there is a navigation bar with buttons: "Create", "Inspect", "Predict", "Login", and "Help". The "Create" button is highlighted. The page content is in a light blue box. It says "This service creates" followed by a list of model types: "lazar classification" and "lazar regression" models. Below this, it states "from your uploaded datasets. Further modelling algorithms will be added in future versions." and "Please read the instructions for creating training datasets before submitting." There is a text input field for "Select training data in Excel or CSV format:" with the value "allvergleich\trmp\hamster_carcinogenicity.csv" and a "Durchsuchen..." button. Below the input field are "Create model" and "Cancel" buttons. At the bottom of the page, it says "© in silico toxicology 2009-2010, powered by OpenTox".

Create and evaluate models to predict toxicity

ToxCreate

Create Inspect Predict Login Help

User: **guest**

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.

Select training data in **Excel** or **CSV** format:

You are currently logged in as **guest** and your models can be modified or deleted by other guests. Please **log in** with your **OpenTox** account to control your model permissions.

© in silico toxicology 2009-2010, powered by **OpenTox**

ToxCreate

ToxCreate - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

TC http://www.toxcreate.org/models

TC ToxCreate

Create and evaluate models to predict toxicity

ToxCreate

Create Inspect Predict Login Help

User: **guest**

Get an overview about ToxCreate models. Parts of this page are refreshed every 5 seconds to update the model status.

Hamster Carcinogenicity ([edit](#))

Status: Completed([delete](#))

Training compounds: 85

Algorithm: [lazar](#)

Type: [classification](#)

Descriptors: [Fminer backbone refinement classes](#)

Training dataset: [Excel sheet](#) , [YAML](#) ([experts](#))

Feature dataset: [Excel sheet](#) , [YAML](#) ([experts](#))

Model: [QMRf Editor](#), [YAML](#) ([experts](#), *models cannot be represented in Excel*)

Validation: [show](#)

Detailed report:

Number of predictions: 69

Correct predictions: 82.68 %

Weighted area under ROC: 0.935

Specificity: 0.143

Sensitivity: 0.865

Confusion Matrix:

		Measured	
		active	inactive
Predicted	active	32	5
	inactive	7	25

© in silico toxicology 2009-2010, powered by [OpenTox](#)

ToxCreate

ToxCreate - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

TC http://www.toxcreate.org/lazar#lazar_algorithm

TC ToxCreate phenylhydrazine (CHEBI:27924)

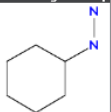
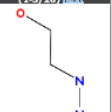
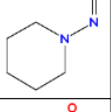
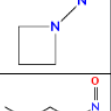
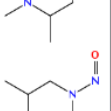
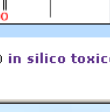
Create and evaluate models to predict toxicity

ToxCreate

Create Inspect **Predict** Login Help

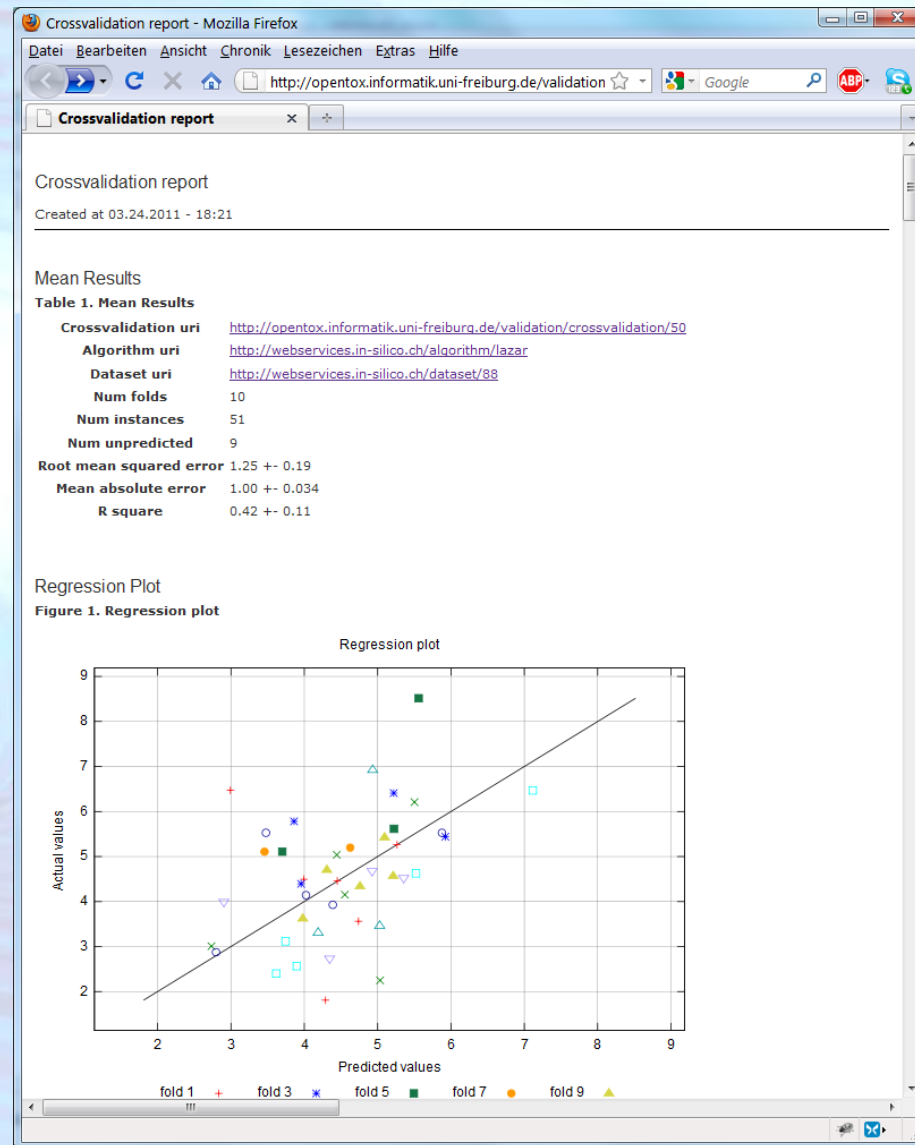
User: guest

New prediction

Hamster Carcinogenicity	Prediction	Confidence	Supporting information
	active	0.108	Names and synonyms Significant fragments
Neighbors (1-5/25) next	Measured activity	Similarity	Supporting information
	inactive	0.715	Names and synonyms Significant fragments
	inactive	0.5	Names and synonyms Significant fragments
	inactive	0.5	Names and synonyms Significant fragments
	inactive	0.5	Names and synonyms Significant fragments
	inactive	0.5	Names and synonyms Significant fragments

© in silico toxicology 2009-2010, powered by OpenTox

ToxCreate



ToxCreate

Crossvalidation report - Mozilla Firefox

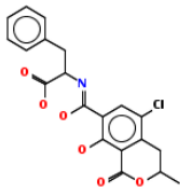
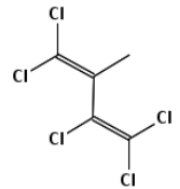
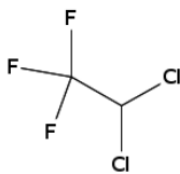
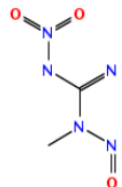

File Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://opentox.informatik.uni-freiburg.de/validation

Crossvalidation report

Predictions

Table 4. Predictions

compound	actual value	predicted value	confidence value	compound-uri
	6.47	2.99	0.41	http://webservices.in-silico.ch/compound/InChI=1S/C20H18ClNO6/c1-10-7-12-14(21)9-13(17(23)16(12)20(27)28-10)18(24)22-15(19(25)/h2-6,9-10,15,23H,7-8H2,1H3,(H,22,24)(H,25,26))
	3.56	4.74	0.47	http://webservices.in-silico.ch/compound/InChI=1S/C5H3Cl5/c1-2(4(7
	1.81	4.29	0.47	http://webservices.in-silico.ch/compound/InChI=1S/C2HCl2F3/c3-1(4
	5.26	5.26	0.41	http://webservices.in-silico.ch/compound/InChI=1S/C2H5N5O3/c1-6(5(H2,3,4))
				

Reporting (Qedit)

Document 1



1. Substance 2. General Information 3. Prediction 4. Adequacy Info



Model Prediction Applicability Domain

Model Information


Model Info. Algorithm Info. Predicted Feature Dependent Feature Download Model Info Find Models Online...



Model & Training Dataset :

Link to Model Resource :  


Link to Training Dataset :  



Training Algorithm :

Algorithm Name : 

Link to Algorithm Resource :  

Predicted Feature :

Predicted Feature Name : 

Link to Feature Resource :  

Related QMRF Report :

QMRF Report (reference) :

QMRF report discussion :

Model Version Info :

Model Version

Model Date

☐ Use current date

Year :

Month :

Day :

Reporting (Qedit)

Qedit application window showing the reporting interface for a compound.

File path: /home/chung/Desktop/phenobarn

Navigation tabs: 1. Substance, 2. General Information, 3. Prediction, 4. Adequacy Info

Subsection: 3.3. Applicability Domain Info.

Form fields:

- Name Applicability Domain Estimation Algorithm Used:
- Link to Applicability Domain Resource:

Subsection: 3.3.b. Structural Analogues

Buttons: Add Compound Wizard, Remove, Clear List, Acquire List of Analogues, Compound Info

Similarity Level:

List of Structural Analogues (URIs):

Chemical Name	Experimental Value
phenobarbital, Phen...	
5-methyl-5-phenylb...	
methylphenobarbit...	
5-allyl-5-phenylbarbi...	
primidone, Primaclo...	
calcium bis[5-(1-cyc...	
5-ethyl-5-(4'-hydrox...	
barbexaclone	
1,3-dimethyl-5-phen...	
5-ethyl-5-phenylbar...	
N-(acetaminophen)...	

Image of structural analogue:

3.3.c. Consideration

Discussion

Applicability Domain Result:

3.3.a. Choose Domain:

Compound Details window:

Compound Details

URI: <http://ambit.uni-plovdiv.bg:8080/ambit2/compound/5100/conformer/5100>

Smiles: [Ca+2].CCC1(C(=O)NC(=NC1=O)[O-])C2=CCCCC2.CCC3(C(=O)NC(=NC3=O)N)C4=CC=CC=C4

InChI:

InChI Key:

CAS number:

Chemical Name:

Einecs:

REACH Reg. Date:

Available Conformers (Links):

- <http://ambit.uni-plovdiv.bg:8080/ambit2/compound/5100/conformer/5100>
- <http://ambit.uni-plovdiv.bg:8080/ambit2/compound/5100/conformer/105301>
- <http://ambit.uni-plovdiv.bg:8080/ambit2/compound/5100/conformer/181274>

Buttons: Close, Apply Changes and Close

Reporting (Qedit)

Qedit application window showing the reporting interface for a compound.

The window title is `/home/chung/Desktop/phenobarn`.

The interface includes tabs for: 1. Substance, 2. General Information, 3. Prediction, 4. Adequacy Info.

Compound Info

Buttons: Save CML, Save RDF, Compound Details, Download Compound Info.

Search for compound (Provide any Keyword like its chemical name, CAS Registration Number, Smiles etc or provide its URI)

Link to Dataset containing descriptors

Compound Name(s) (Synonyms):

- phenobarbital, Phenobarbital, Phenobarbitol, BARBE
- PHENOBACA
- PHENOBAMG
- PHENOBARB

Buttons: + Add Synonym, - Remove Synonym, Clear All

Structure Image

Chemical structure image of Phenobarbital (5-ethyl-5-phenyl-1,3-dimethyl-2,4,6-trioxo-1,2,3,4-tetrahydropyrimidin-2-ylidenehydrazine).

Descriptors

Buttons: +, -, Erase, Copy

Descriptor	Value
------------	-------

Lock your report

You can lock your report so that it will not be opened by the editor unless a passphrase is provided. However this should be considered a low strength security measure!

Passphrase:

Buttons: OK, Cancel

Features of the Substance

Chemical features that may affect predictions for the substance

Model resource

Dataset Resource

Algorithm
service

Feature service

Descriptor resource

Assay resource

Chemical compound

Feature service

Compound service

Linked resources: Compound, Algorithm, Model, Dataset, Features

Dataset
Resource

Descriptor
resource

Assay
resource

Chemical
compound

Blue Obelisk
algorithms
ontology

Regression
Classification
Chemistry
Descriptors etc.

OpenTox
algorithm types
ontology

Toxicology related
ontologies

http://apps.ideaconsult.net:8080/ambit2/dataset/R545

data Entry	compound
values	http://apps.ideaconsult.net:8080/ambit2/compound/38/conformer/419609
feature	TopoPSA
Feature value	6.480000019073486
type	Feature Value
feature	nHBDon
Feature value	0.0
type	Feature Value
feature	caco2
Feature value	8.849999904632568
type	Feature Value
feature	WNSA-3
Feature value	-374800205230713
type	Feature Value
feature	FPSA-2
Feature value	0.8797000050544739
type	Feature Value
compound	http://apps.ideaconsult.net:8080/ambit2/compound/144824/conformer/419615
values	TopoPSA
feature	210.5399938861328
Feature value	Feature Value
feature	nHBDon
Feature value	5.0
type	Feature Value
feature	caco2
Feature value	-5.920000076293945
type	Feature Value
feature	WNSA-3
Feature value	-64.12879943847656
type	Feature Value
feature	FPSA-2
Feature value	2.147799968719482
type	Feature Value

http://apps.ideaconsult.net:8080/ambit2/feature/22213

Name of the algorithm	type	Class
type	type	Class
Numeric Feature	type	Class
Source	type	subClassOf
Units	type	ObjectProperty
nHBDon	type	DatatypeProperty
Source	sameAs	http://www.blueobelisk.org/ontologies/chemoinformatics-algorithms/#nBondDonors
Title		nHBDon
Source		http://apps.ideaconsult.net:8080/ambit2/algorithm/org.openscience.cdk.qsar.descriptors.molecular.HBondDonorCountDescr
Units		
type		Numeric Feature

http://apps.ideaconsult.net:8080/ambit2/feature/22200

Numeric Feature	type	Class
type	type	Class
Source	type	subClassOf
Units	type	ObjectProperty
caco2	type	DatatypeProperty
Source		caco2
Units		c049084m_caco2-training_set.sdf
type		Numeric Feature
sameAs		Gastrointestinal absorption

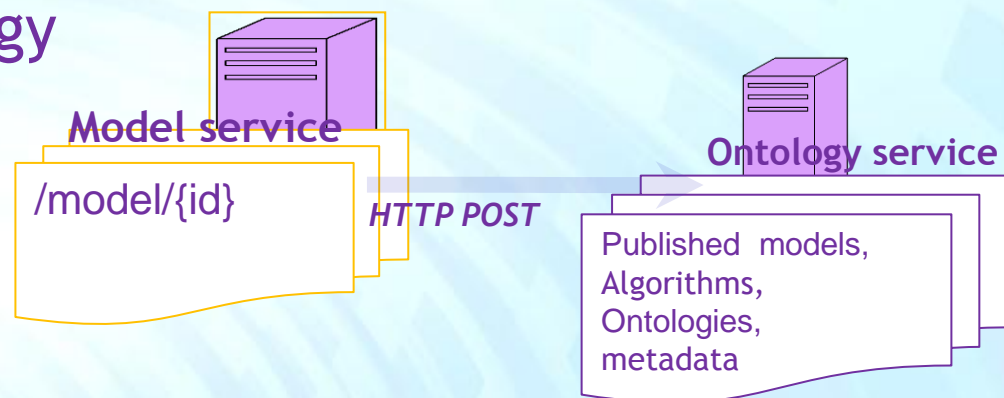
Make the model available

Register at OpenTox ontology service

- RDF triple storage
- Accepts HTTP POST
- SPARQL endpoint

Curl -X POST -d
"uri=http://apps.ideaconsult.
net:8080/ambit2/model/57"
<http://apps.ideaconsult.net:8080/ontology>

Becomes visible for applications



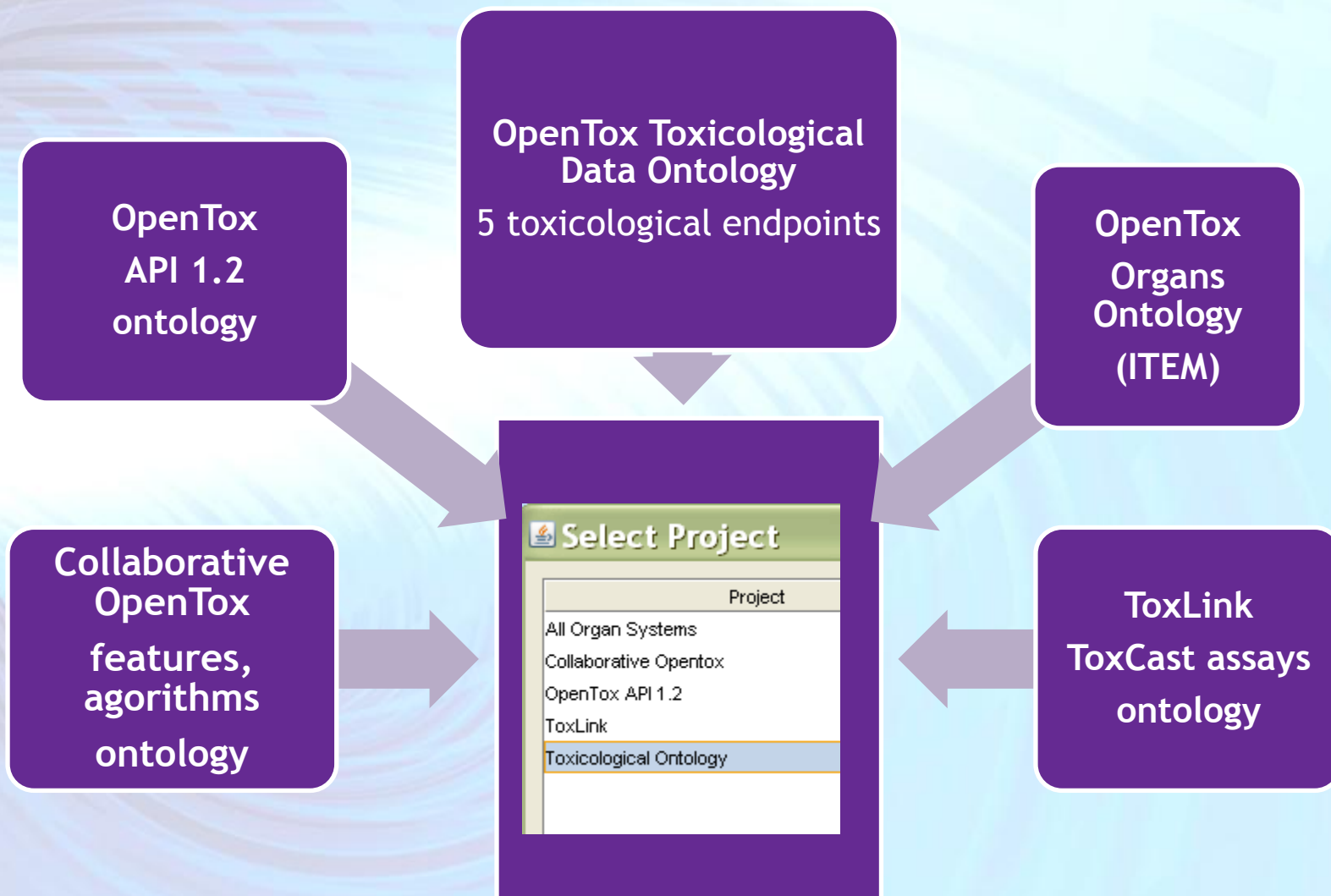
Model	Endpoint	Algorithms	Validation
CholesterolWeight	Acute toxicity to fish (pH ₅₀)	MolecularWeight	
ToxTree: Verhaar scheme for predicting toxicity mode of action	ToxTree: Verhaar scheme for predicting toxicity mode of action	ToxTree: Verhaar scheme for predicting toxicity mode of action	
ToxTree: Benign/Boss rules for carcinogenicity and mutagenicity	Carcinogenicity	ToxTree: Benign/Boss rules for carcinogenicity and mutagenicity	
EqKa	Dissociation constant (pKa)	pKa	
ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	
ToxTree: Michael acceptors	Endpoints	ToxTree: Michael acceptors	
ToxTree: Eye Irritation	Eye irritation/corrosion	ToxTree: Eye Irritation	
ECaco-2 Cell Permeability	Gastrointestinal absorption	Regression: Linear regression	Model validation report
OpenTox model created with TUM's ML Regression model learning web service	Gastrointestinal absorption	http://openTox.informatica.uni-muenchen.de/8080/OpenTox-dev/algos/MLRegression	
OpenTox model created with TUM's ML Regression model learning web service	Gastrointestinal absorption	http://openTox.informatica.uni-muenchen.de/8080/OpenTox-dev/algos/MLRegression	
EqKa	Human health effects	Lipinski Rule of Five	
ToxTree: Cramer rules	Human health effects	ToxTree: Cramer rules	
EqKa	Human health effects	EqKa	

Need for communications in the community overcoming different languages and vocabularies



Explaining the rules of different games on a
conservation project trip in the Caprivi, Namibia

Collaborative Ontology Development: Collaborative Protege Server



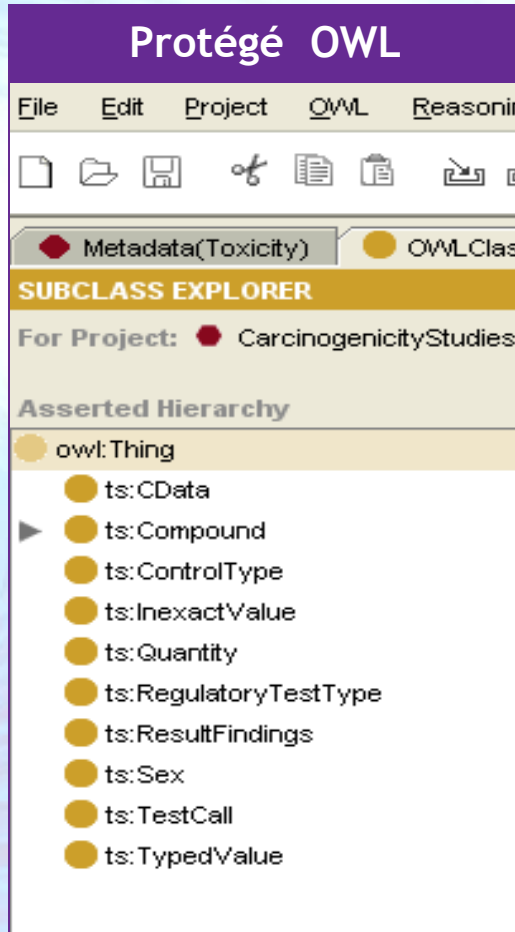
ToxML: conversion in OWL ontology

- Initial work to check out how costly would be to convert ToxML to an OWL ontology
- Why:
 - -to integrate the Leadscope databases in the OpenTox service
 - -to combine with data coming from different databases (e.g. for complex queries)

ToxML
xsd
schema



Protégé OWL



Some parts of the
taxonomy may need to
be reorganized

Procedures:

➤ each element which doesn't have a type in the schema is converted to an OWL class

each element which has a type in the summary file is considered as a property

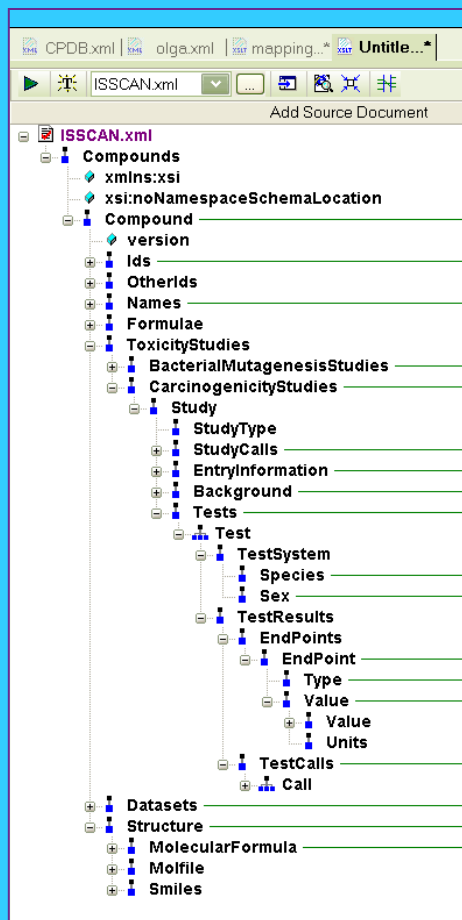
the parent relationships are kept



Needs for extensions: e.g. target sites, target cells, species are free text fields.
Solutions:
development extension in OWL, e.g. Organs Ontology has been developed importing of parts of the neighboring ontologies

Toxicological Endpoint Ontology Development

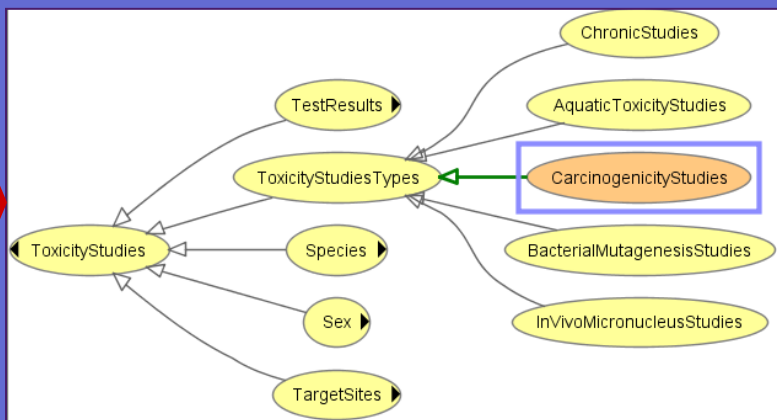
ToxML schema



Other publicly available resources:
DSSTox, GoReni (ITEM), ISSCAN ...

OpenTox
Toxicological
Endpoint
Ontology

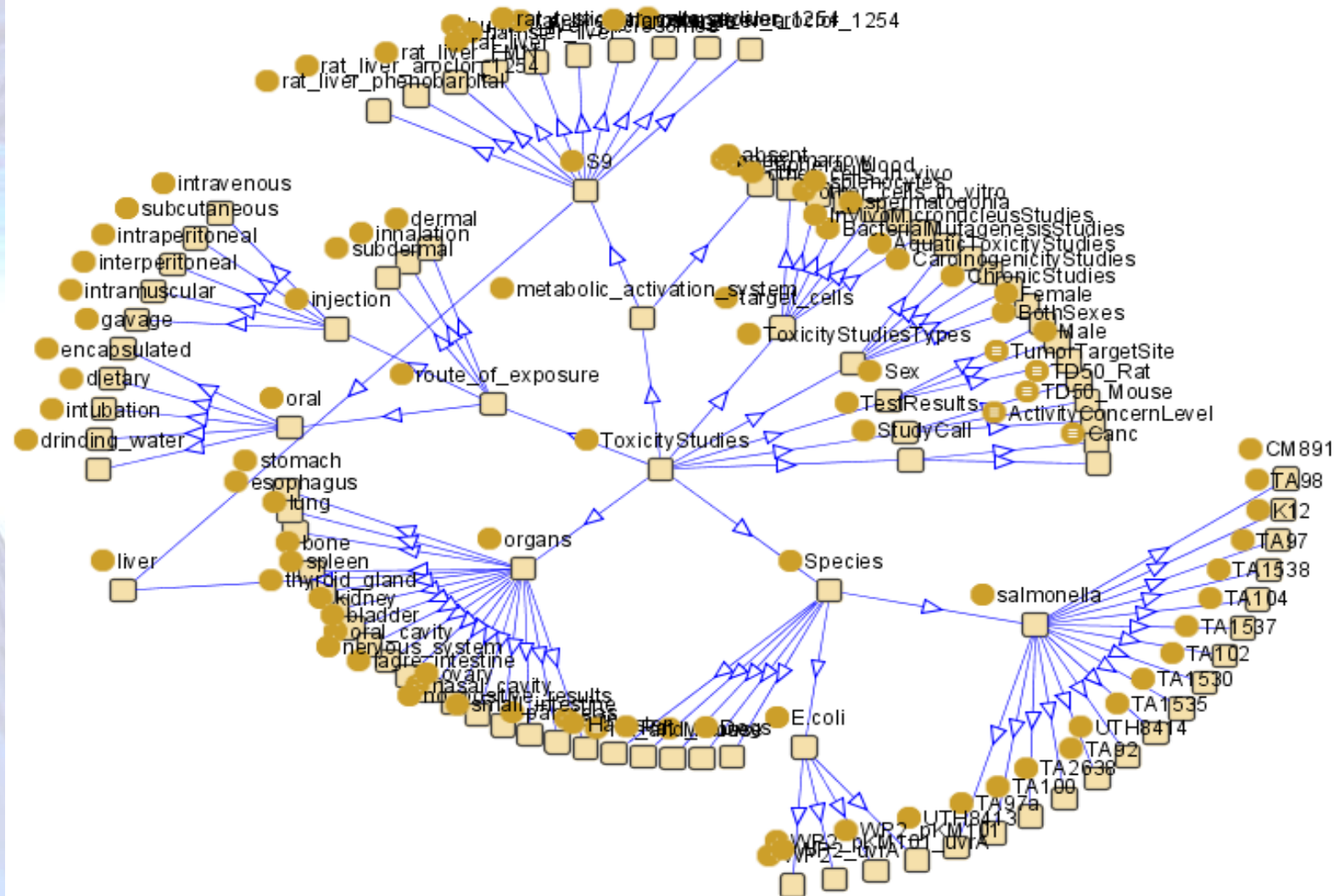
Ontology Development



Re-use of terms defined in
neighbouring ontologies (e.g. OBO)

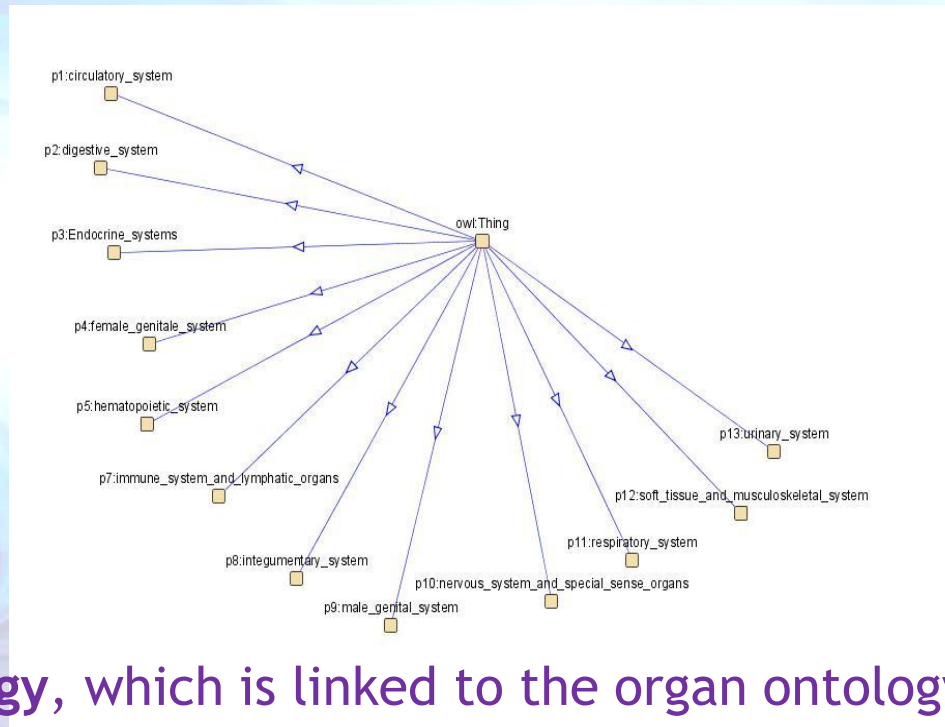
Collaborative
Protégé
Environment

Toxicological Ontology: graphical representation



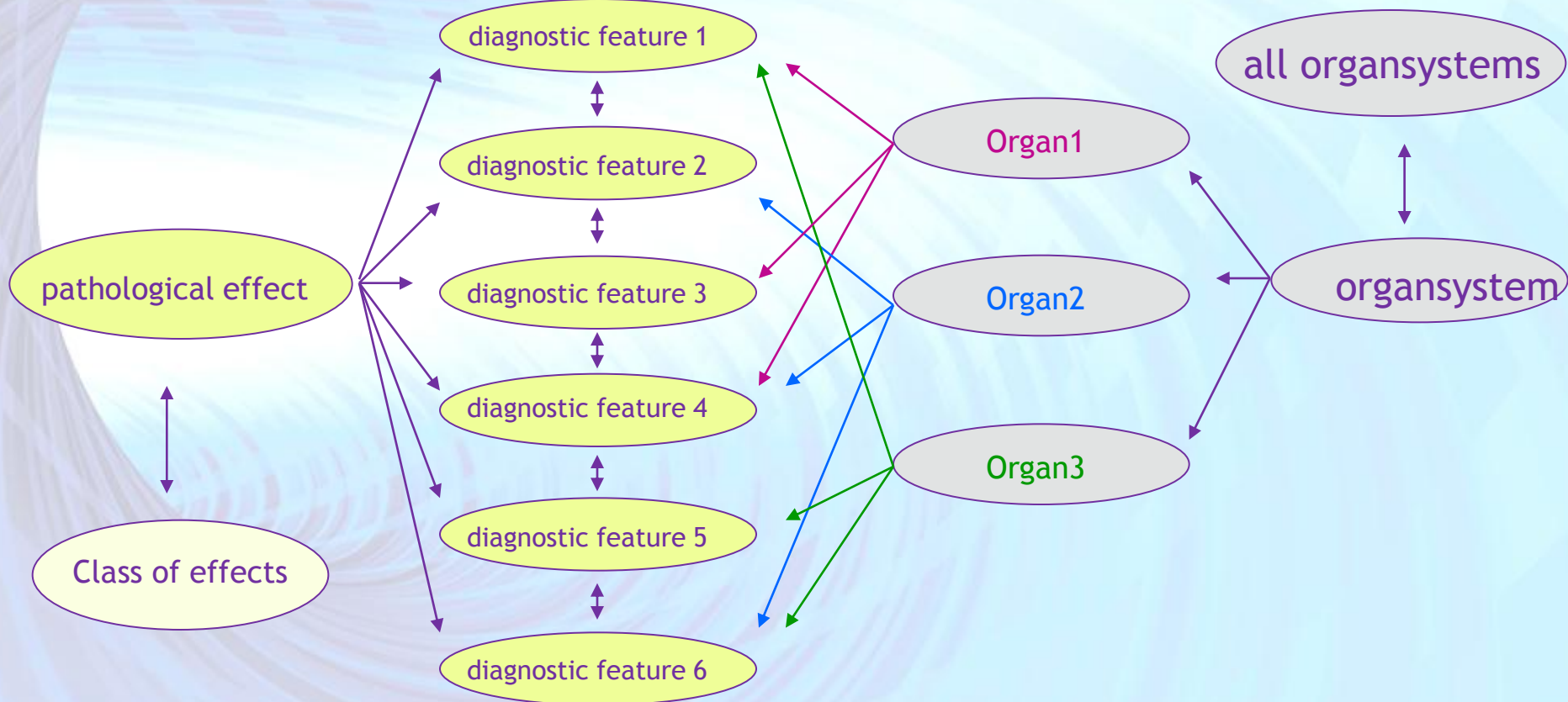
OpenTox Organ Ontology Development

- organ ontology consisting of 12 very detailed organ systems



- effect ontology, which is linked to the organ ontology
- comprehensive review by FhG pathologists, who have been involved in the INHAND process

OpenTox Organ Ontology

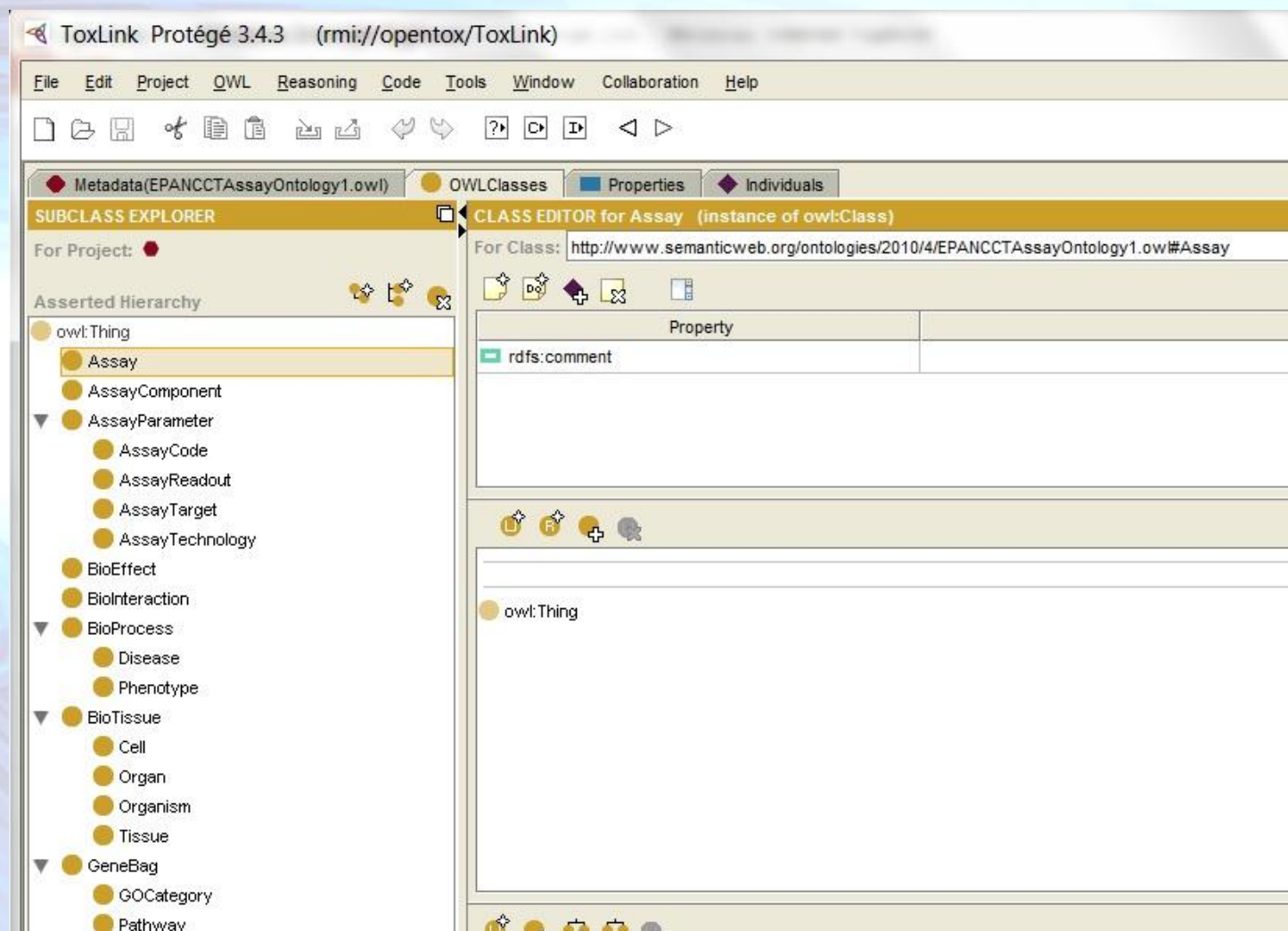


A Toxicology Ontology Roadmap

In Preparation by: Barry Hardy (Douglas Connect and OpenTox), Ian Dix (AstraZeneca & Pistoia Alliance), Sherri Matis-Mitchell (AstraZeneca), David Cook (AstraZeneca), David Heard (Novartis), Dominic Clark (EMBL-EBI), John Overington (EMBL-EBI), Philip Judson (Lhasa), David Watson (Lhasa), Anne Hersey (EMBL-EBI), Andrew White (Unilever), Loca Toldo (Merck KGaA), Gordana Apic (Cambridge Cell Networks), Imran Shah (US EPA), Chihae Yang (Altamira), Dave Bower (Leadscope), Ola Spjuth (Univ Uppsala), Janna Hastings (EMBL-EBI), Philip Carthew (Unilever), ----

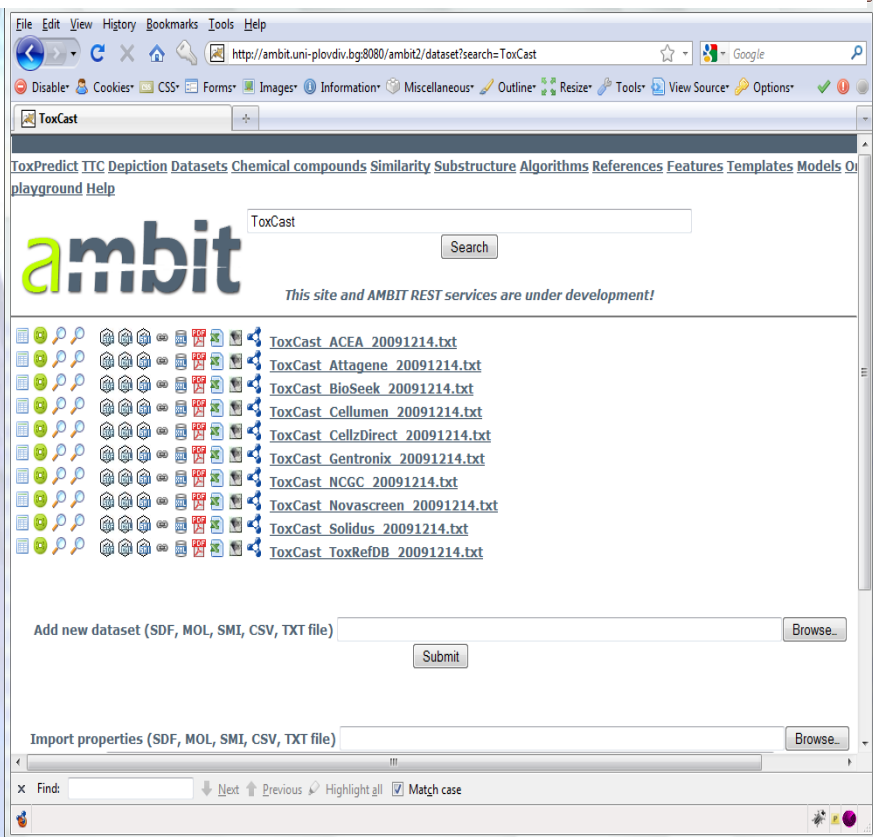
Based on Proceedings from the Toxicology Ontology Roadmap Workshop
EMBL-EBI Industry Programme Workshop
16 -17th November 2010, Hinxton, UK

ToxLink: ToxCast Ontology



Example: ToxCast

Dataset service at
<http://ambit.uni-plovdiv.bg:8080/ambit2/dataset>



Query an OpenTox ontology service at
<http://ambit.uni-plovdiv.bg:8082/ontology>

```
PREFIX ot:<http://www.opentox.org/api/1.1#>
PREFIX ota:<http://www.opentox.org/algorithms.owl#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX otee:<http://www.opentox.org/echaEndpoints.owl#>
PREFIX toxcast:<http://www.opentox.org/toxcast#>

select *
where {
  ?Feature rdf:type ot:Feature.
  {?Feature dc:title ?title}.
  {?Feature owl:sameAs ?assay}.
  {?assay toxcast:gene ?geneid}.
  {?assay toxcast:hasProperty ?species}.
  {?species rdfs:type toxcast:SPECIES}.
  {?assay toxcast:hasProperty ?target_source}.
  {?target_source rdfs:type toxcast:ASSAY_TARGET_SOURCE}.
  {?assay toxcast:hasProperty ?target_family}.
  {?target_family rdfs:type toxcast:ASSAY_TARGET_FAMILY}.
  {?assay toxcast:hasProperty ?target}.
  {?target rdfs:type toxcast:ASSAY_TARGET}.
  {?assay toxcast:hasProperty toxcast:Cytochrome_P450}.
}
order by ?feature ?assay ?target
```

Example: ToxCast

PREFIX ot:<<http://www.opentox.org/>
 PREFIX ota:<<http://www.opentox.org/>
 PREFIX owl:<<http://www.w3.org/2002/>
 PREFIX dc:<<http://purl.org/dc/element/>
 PREFIX rdfs:<<http://www.w3.org/2000/>
 PREFIX rdf:<<http://www.w3.org/1999/>
 PREFIX otee:<<http://www.opentox.org/>
 PREFIX toxcast:<<http://www.opentox.org/>

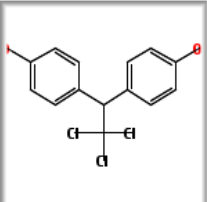
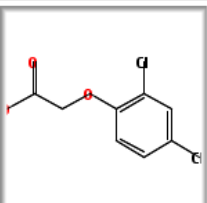
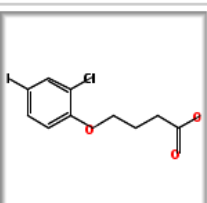

```

    select ?Feature ?title ?id
    where {
        ?Feature rdf:type ot:Feature
        {?Feature dc:title ?title}.
        {?Feature owl:sameAs ?assay}.
        {?assay toxcast:gene ?geneid}.
        {?assay toxcast:hasProperty ?genename}.
        {?genename rdf:type toxcast:GENE_1}
    }
    
```

Query an OpenTox ontology service
<http://ambit.uni-plovdiv.bg:8082/>

Chemical compounds

Search results Dataset = 961 hits: 100

#	Compound	ToxCast At	Benigni /	Benigni /
		ATG RORE CIS	Structural Alert for genotoxic carcinogenicity	Structural Alert for nongenotoxic carcinogenicity
1		1000000.0	NO	NO
2		1000000.0	NO	NO
3		1000000.0	NO	NO
4		1000000.0	NO	NO

[dataset/961?feature/335126](#)

OpenToxipedia



Barry Hardy Log out Quicktools Site Setup Help

Site Map Accessibility Contact Data

Search Site

Home Toxicity Prediction OpenTox Blog People Partners Development OpenToxipedia

User Guidance Latest Entries A B C D E F G H I J K L M N O P Q R S T U V W
X Y Z by Categories Entries OpenToxipedia

You are here: Home » OpenToxipedia

Contents View Edit Rules Sharing History

Actions Display Add new... State: Published

OpenToxipedia

by Barry Hardy — last modified Sep 03, 2009 01:09 PM

OpenTox Community Resource for Toxicology Vocabulary and Ontology

OpenTox is supporting the creation and curation of OpenToxipedia, a community-based predictive toxicology knowledge resource. All members of the community are welcome to provide entries, suggested definition edits or additional information to entries in the resource.

OpenTox is supporting the application and development of the **ToxML** standard for representation of toxicology data, the **OECD principles for (Q)SAR model validation**, and the use of the **OECD HT** standard for regulatory reporting purposes.

OpenToxipedia provides here a Vocabulary Resource of toxicology terminology. We hope you find the resource useful and consider contributing to terms and their content.

Guidance for Vocabulary Resource entries



www.opentox.org/opentoxipedia

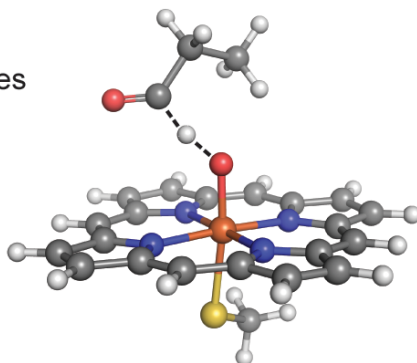


SMARTCyp Service for Predicting Metabolites

Atom Reactivity Library

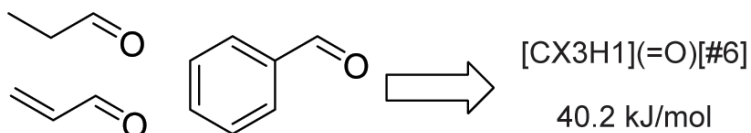
A. Calculate Quantum Chemical Reference Energies

Calculate transition state energies using density functional theory



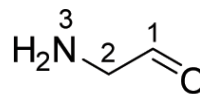
B. Define SMARTS Rules

Group calculations by fragments and calculate average energies



SMARTCyp

1. Assign Energies By SMARTS matching



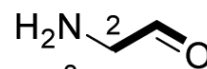
Atom	SMARTS	Energy
1	[CX3H1](=O)[#6]	40.2
2	[CX4][N]	39.8
3	[N^3][H1,H2]	54.1

2. Compute Accessibility Descriptor

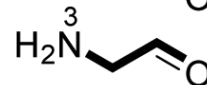
$$A_i = \text{Maxbonds}_i / \text{Maxbonds}_{\text{all}}$$



$$A_1 = 2 / 3 = 0.67$$



$$A_2 = 2 / 3 = 0.67$$



$$A_3 = 3 / 3 = 1.00$$

3. Compute Score and Rank Atoms

$$\text{Score, } S = E - 8A$$

Lowest score gets rank 1

$$S_1 = 40.2 - 8 \cdot 0.67 = 34.84$$

$$S_2 = 39.8 - 8 \cdot 0.67 = 34.44$$

$$S_3 = 54.1 - 8 \cdot 1.00 = 46.10$$

Atom 1 - Rank 2

Atom 2 - Rank 1

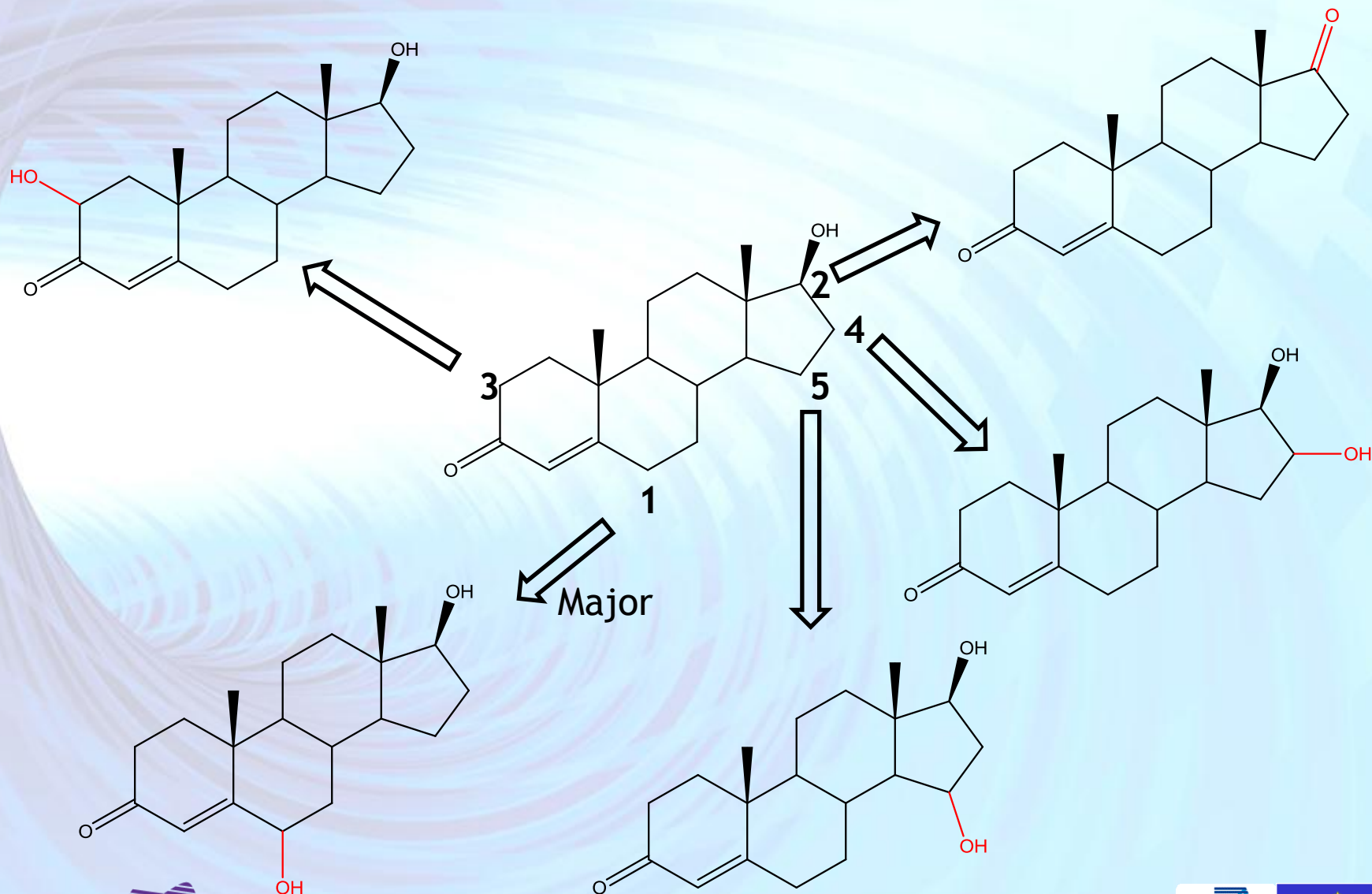
Atom 3 - Rank 3



SMARTCyp - developed by Patrik Rydberg, University of Copenhagen

www.farma.ku.dk/index.php/SMARTCyp/7990/0/

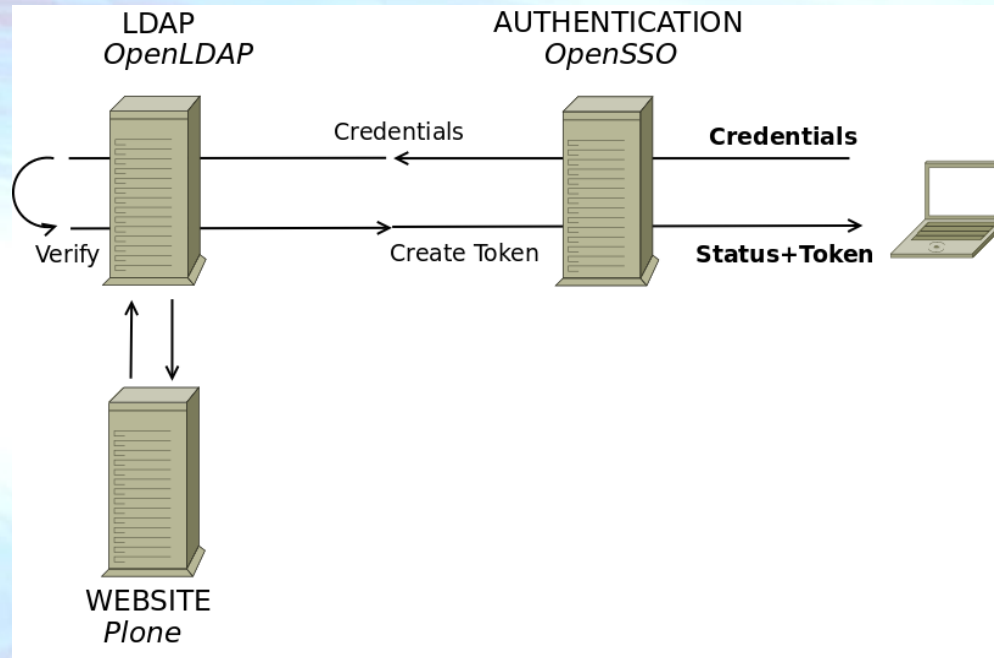
SmartCYP Prediction of Testosterone Metabolites



Controlling Access to Confidential Information

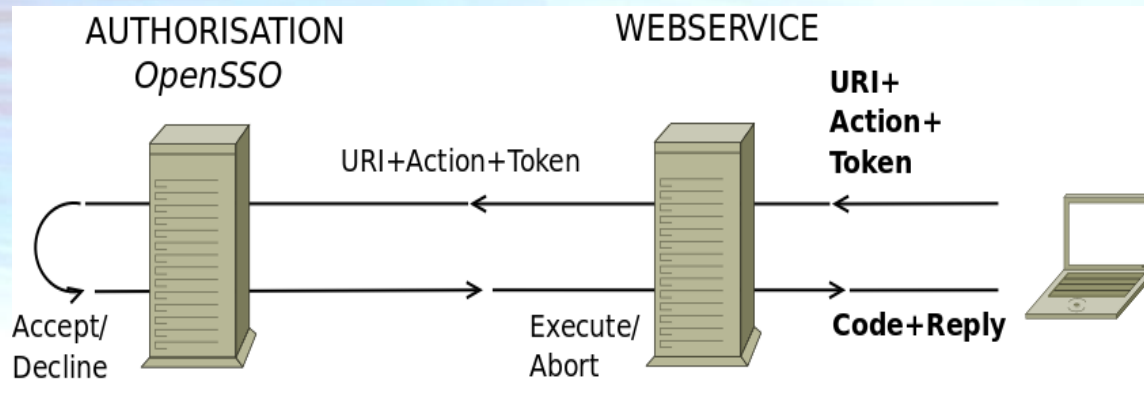
- OpenTox makes resources available through **URIs**
- OpenTox provides facilities to protect confidential information located at **URIs**. Two tasks are involved here:
 - **Authentication**: Confirming the identity of the user requesting access
 - **Authorisation**: Granting the confirmed identity access according to a set of restrictions described in policies

Authentication



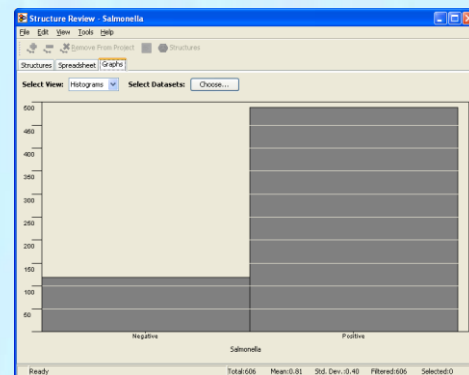
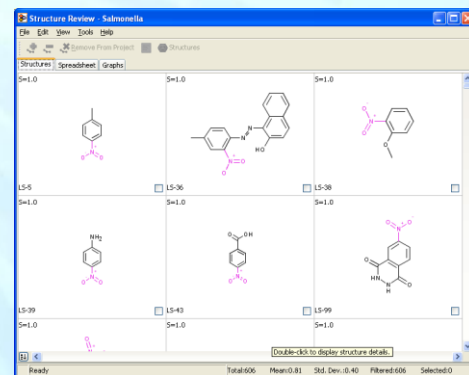
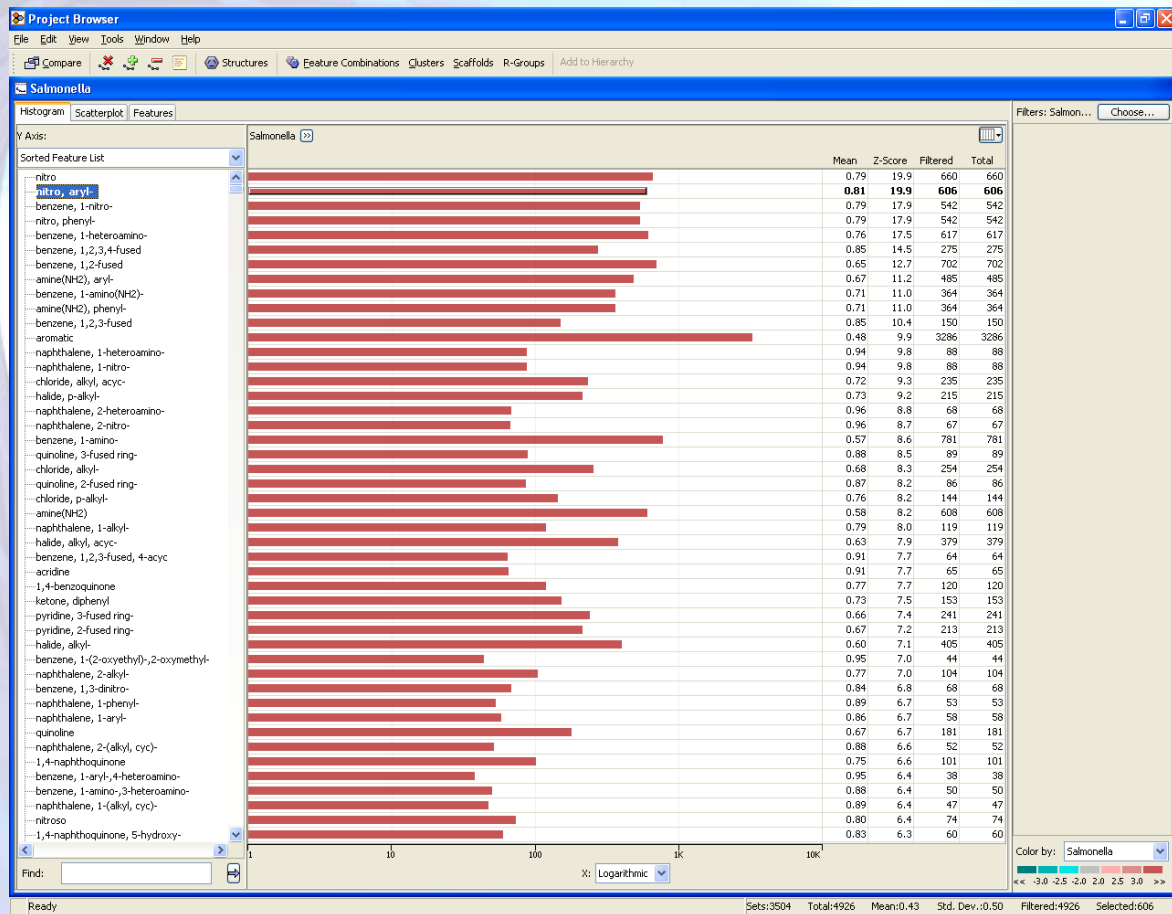
- Registered users are instantly available as potential users of OpenTox web services
- Users receive a token upon service request

Authorisation



- Tokens encode user identity
- Tokens are valid for a certain time period only (customizable)
- The triplet URI+Action+Token makes up the call to be authorised
- All messages are encrypted (SSL)
- Resource Owners create and modify policies defining access rules

OpenTox - Leadscope





UPPSALA
UNIVERSITET

Problem

Building
Blocks

Conclusion

The Chemistry Development Kit

A Family of Projects

- CDK-Taverna (chemoinformatics workflows)
- JChemPaint (semantic 2D editor)
- ChemoJava (GPL-ed extension)

Goals

- library of cheminformatics algorithms
- educational

Usage

- CDK: 100+ times cited in scientific literature
- Bioclipse, KNIME, Jumbo (CML), AMBIT, ...

C. Steinbeck et al., J.Chem.Inf.Comput.Sci, 2003

C. Steinbeck et al., Curr.Pharm.Design, 2006

2010-05-30

Bioclipse & Proteochemometric Group

- 9 -

Egon Willighagen | chem-bla-ics.blogspot.com

OpenTox - Bioclipse



UPPSALA
UNIVERSITET

Problem

Building
Blocks

Conclusion

Bioclipse

The screenshot displays the Bioclipse application window. On the left is the 'Bioclipse Navigator' pane showing a tree view of files, with 'starfile31.sdf' selected. The main workspace contains a table with two rows of chemical structures. The first row has a molecular weight of 14577 and a molecular mass of 17159. The second row has a molecular weight of 14570 and a molecular mass of 17160. On the right is the 'Properties' pane, which shows the 'General' properties for the selected structure, including 'Has 2D Coords' (yes), 'Has 3D Coords' (no), 'Molecular Format' (N/A), 'Molecular Formula' (C18H26N6O6P), 'Molecular Mass' (358.1781), and 'Molecular Properties' (MOLREGNO: 17159).

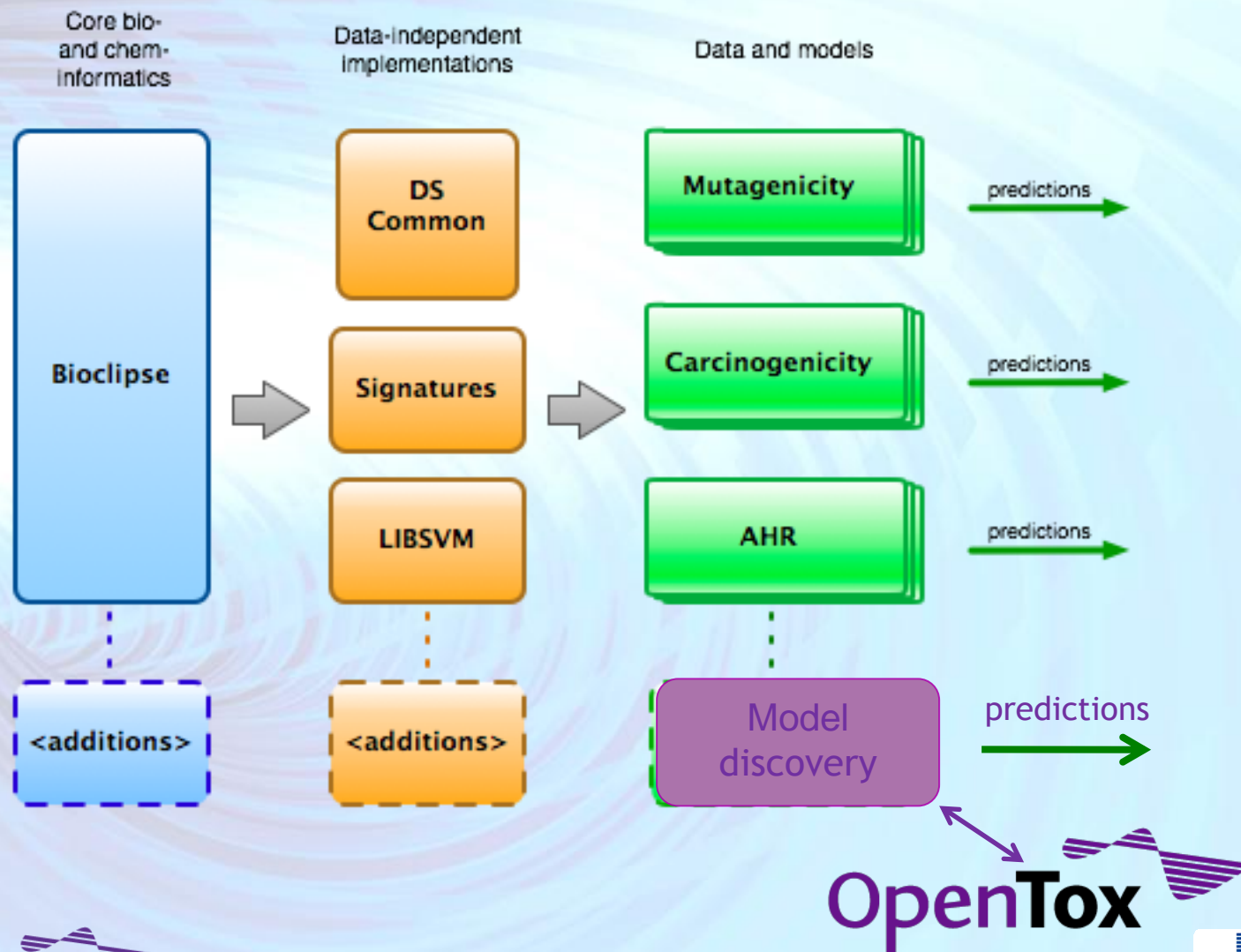
	2D-structure	MOLREGNO
14577		17159
14570		17160

Properties: **General**

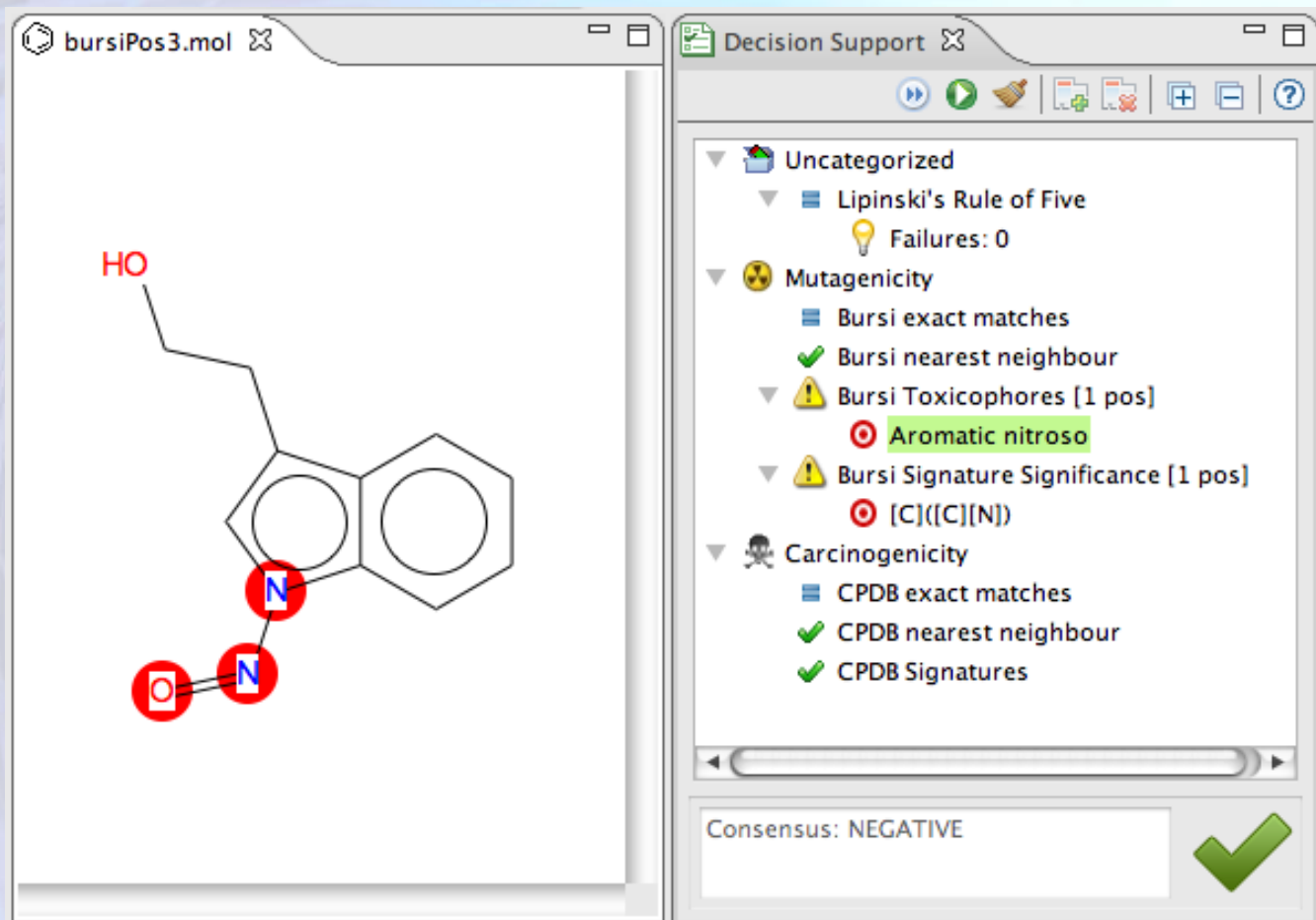
Property	value
Has 2D Coords	yes
Has 3D Coords	no
Molecular Format	N/A
Molecular Formula	C18H26N6O6P
Molecular Mass	358.1781
Molecular Properties	
MOLREGNO	17159

O. Spjuth et al., BMC Bioinformatics 2007, 8:59

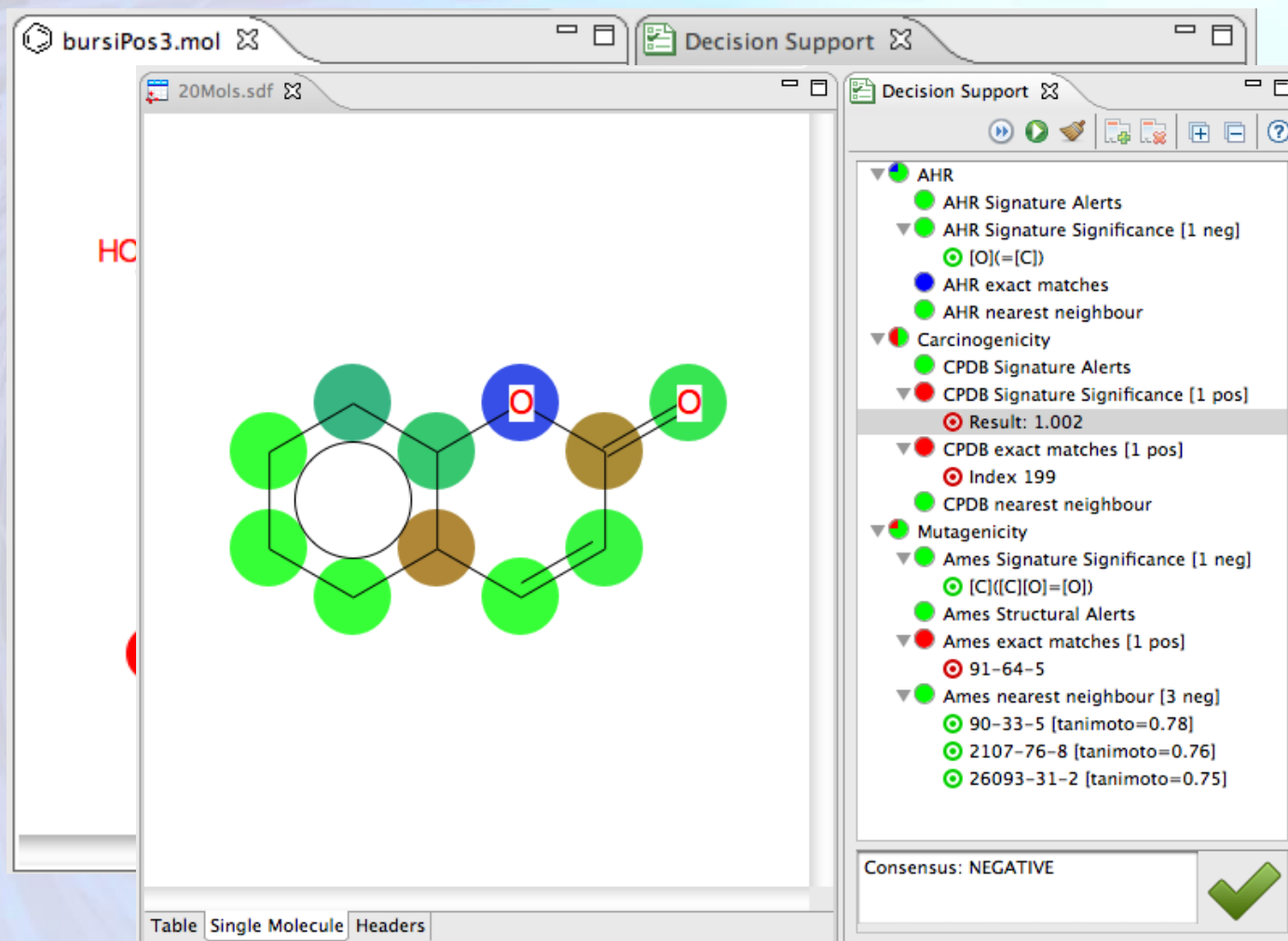
Bioclipse - OpenTox Interoperation



Bioclipse Visualisation Workbench



Bioclipse Visualisation Workbench



Bioclipse Visualisation Workbench - OpenTox

HC

The screenshot displays the Bioclipse Visualisation Workbench interface. The main window shows a chemical structure of a steroid-like molecule with a red epoxide group. The word "Changed" is written in purple above the structure. The interface includes several panels:

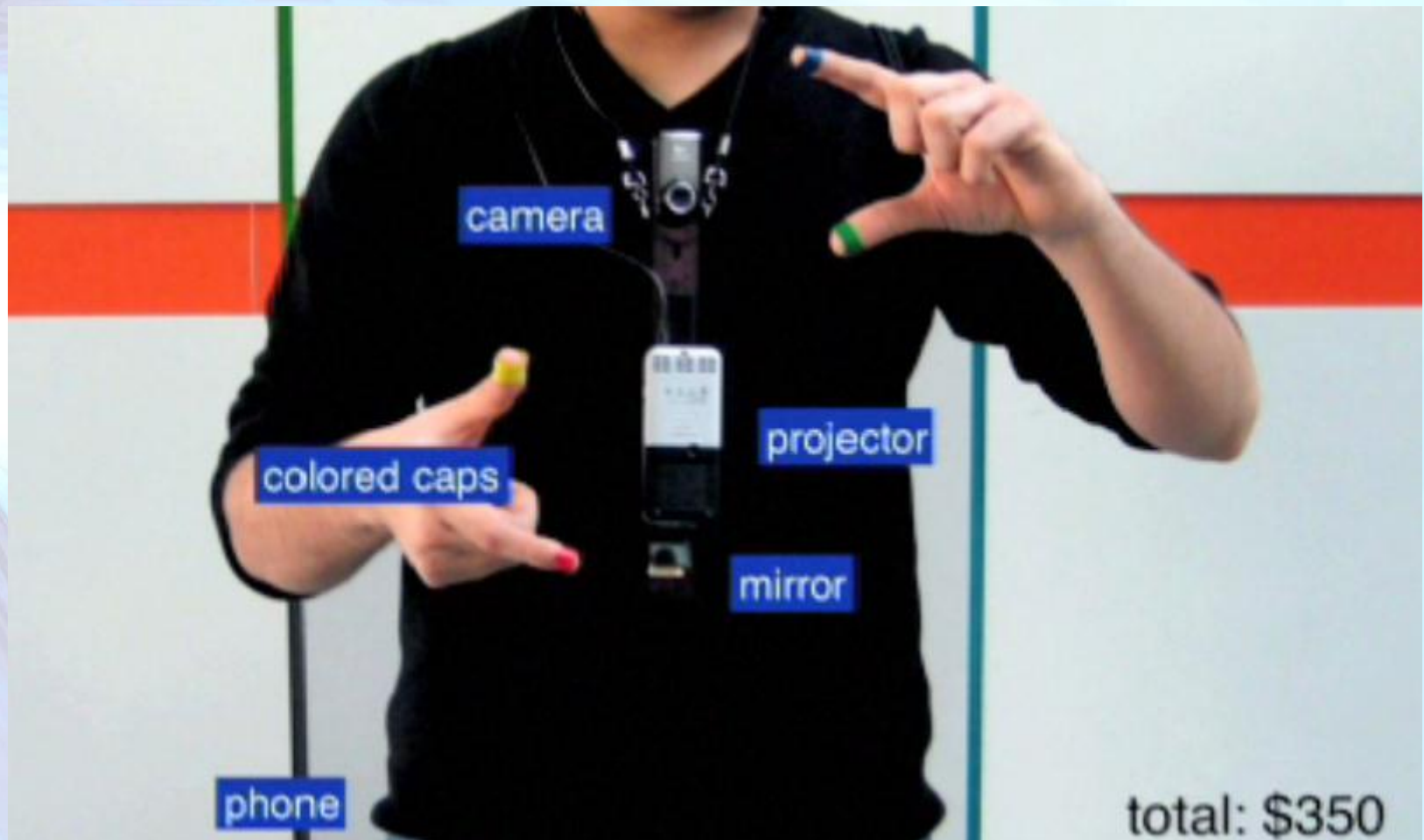
- Decision Support** (top right):
 - Ames Structural Alerts [1 pos]
 - Epoxide
 - Ames exact matches [no hits]
 - Ames nearest neighbour [3 pos, 1 neg]
 - 26761-45-5 [tanimoto=0.82]
 - 2461-18-9 [tanimoto=0.81]
 - 2461-15-6 [tanimoto=0.73]
 - 5926-90-9 [tanimoto=0.71]
 - OpenTox
 - Caco-2 Cell Permeability <http://www.n>
caco2 = -4.548099994659424
 - Lipinski Rule of Five
LipinskiFailures = 0.0
 - MolecularWeight
- Properties** (bottom left):

Property	Value
General	
Classification	POSITIVE
Matching atoms	22, 21, 23
Name	Epoxide
Test	Ames Structural Alerts
- 2D-Structure** (bottom right):

The “Tamboti Tree” Use case



Augmented Reality



Processing Packaging Information



MIT Media Lab

*Inspiration for Saturday Shopping Do
the right thing for Safety Use Case*

OpenTox - ToxCast

U.S. ENVIRONMENTAL PROTECTION AGENCY

ACToR: Aggregated Computational Toxicology Resource

Search: All EPA: This Area:

Data Collection: EPA CCL3

Name: EPA CCL3 List out

Description: EPA has drinking water regulations for more than 10 contaminants. The Safe Drinking Water Act (SDWA) includes a process that we must follow to identify and set unregulated contaminants which may require a national drinking water regulation in the future. EPA must periodically publish the list of contaminants called the Contaminant Candidate List or CCL.

ID: 139

Institutional Source: EPA



Source Type: Chemicals

Number of Substances: 93

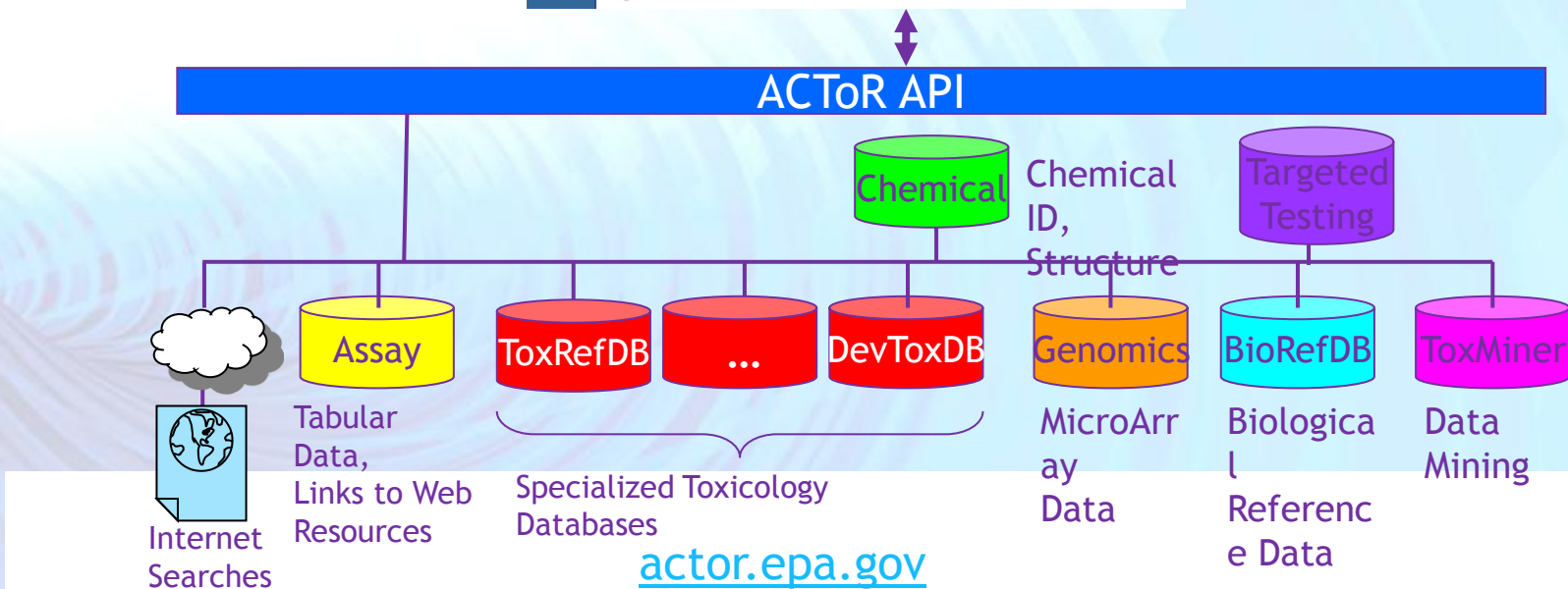
Number of Genetic Chemicals: 92

Chemical Table

Page 1 of 2

Structure	Name	CASRN	Generic Chemical Details	Health	Chemical	Genetics	Developmental	Reproductive	Neurological	Immunological	Endocrine	Other
	1,1,1,2-Tetrachloroethane	630-20-6	Details	H	C	A	S	R	D			
	1,1-Dichloroethane	75-34-3	Details	H	C	A	S	R	D			

ACToR Web Browser



OpenTox - ToxCast

ToxPredict

OpenTox demo application

Step 1: Search
Select structure(s)

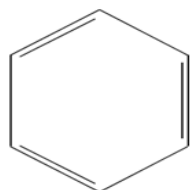
Step 2: Verify structure
Verify structure

Step 3: Models
Select prediction models

Step 4: Estimate
Estimate

Step 5: Results
Display results


This page lists your ToxPredict workflow results for the structure(s) you have selected and the model prediction(s) you have chosen to run. You could also retrieve the ToxPredict report in various other formats, e.g. **SDF, CML, SMI, PDF, CSV, ARFF, RDF/XML or RDF/N3**.




CAS RN
EINECS
IUPAC name
Synonym

71-43-2
200-753-7
benzene
(6)annulene; benzine; Benzol; Benzolene;
bicarburet of hydrogen; carbon oil; Coal naphtha;
cyclohexatriene; mineral naphtha; motor benzol;
nitration benzene; Phene; Phenyl hydride;
pyrobenzol.
21742.0
Benzene
benzene
OK

Synonym
Synonym
Synonym
Quality label

MolecularWeight  **MolecularWeight**
MW

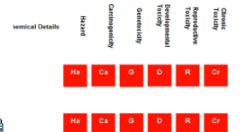
78.1112

Download as 

Welcome, **guest**
Help
Admin

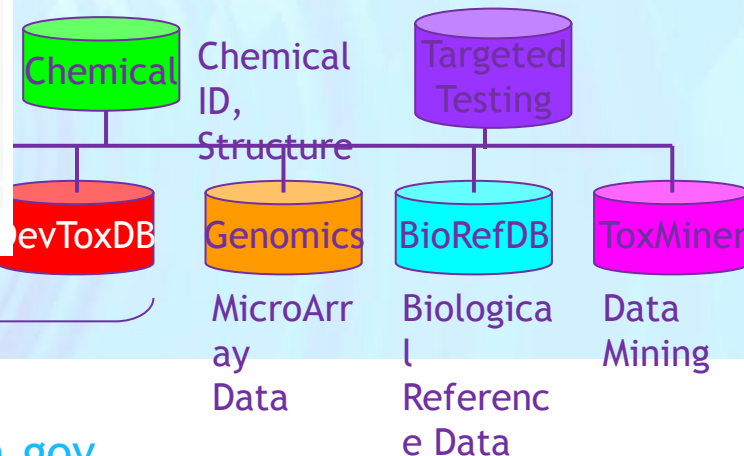
U.S. ENVIRONMENTAL PROTECTION AGENCY
Resource

The Safe Drinking Water Act (SDWA) includes a process that we must follow to identify and list unregulated chemicals in the Nation. EPA must periodically publish this list of contaminants (called the Contaminant Candidate List or CCL).



ACToR Web
Browser

OR API



Chemical Space Visualisation (Ches-Mapper)



Ches-Mapper: Chemical Space Mapping and Visualization in 3D

<http://opentox.informatik.uni-freiburg.de/ches-mapper>



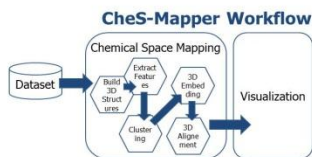
Martin Gütlein^{1*}, Andreas Karwath¹, Stefan Kramer²

*guetlein@informatik.uni-freiburg.de

¹Institute for Computer Science • Albert-Ludwigs-Universität Freiburg • Germany, ²Institute for Computer Science I12 • Technische Universität München • Germany

Abstract

Scientific researchers in the field of cheminformatics, are often overwhelmed by the size and the sheer complexity of chemical datasets. Therefore, the need for visualization tools, is one of the utmost requests. Our recently developed 3D molecular viewer Ches-Mapper (Chemical Space Mapper) includes many techniques, like state-of-the-art structural clustering, and multi-dimensional embedding techniques. Large datasets are divided into clusters of similar compounds and consequently arranged in 3D space, such that their spatial proximity reflects their chemical similarity. This intuitively provides essential information to the user, while making the dataset more easily accessible and allowing easy and understandable access to a large number of chemical structures within seconds. The different clustering approaches employed in our tool utilize common substructures as well as quantitative chemical descriptors of the compounds. These features can be highlighted within Ches-Mapper, which aids the chemist to better understand the underlying scientific knowledge. As a final function, the tools can also be used to select and export specific part of a given dataset for further analysis.



Wizard Dialog to Control Mapping

- A wizard dialog guides through the Mapping process
- Suitable for novice and expert users
- Single Steps:
 - Load dataset
 - Build 3D structure
 - Extract features
 - Clustering
 - 3D Embedding
 - 3D Alignment
- Automatic detection and plug in of new methods and algorithms



Chemical Space Mapping

Build 3D Structure and Extract Features

- Select input dataset
 - Various dataset formats are supported (csv/json/indirect...)
 - Dataset can be directly loaded from the web
- 3D structure is built
 - 3D structure can be built with Chemical Development Kit (CDK) or OpenBabel
 - External libraries like Corina can be plugged in easily
- Extract features
 - Features are required for clustering and embedding
 - Automatic extraction of dozens of descriptors with CDK



Feature	Value
LogP	2.4101
MolWeight	133.0708
MolWeight	133.0708
MolWeight	133.0708
MolWeight	133.0708

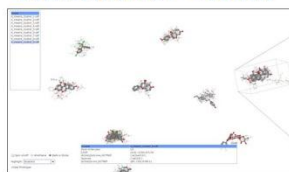
3D Embedding (of Clusters & Compounds)

- Embedding algorithms assign 3D coordinates to each compound or cluster, according to the feature values of the compounds
- Different approaches are provided:
 - Principal Component Analysis (PCA)
 - Multi-dimensional Scaling Using Majorization (SMACOF)
 - T-distributed Stochastic Neighbor Embedding (tSNE)
- Developers can easily plug in their own preferred 3D Embedding algorithm

Features	LogP	MolWeight	Wt path number	Wt polarity number
CCCCCCCCCCCCCCCC	4.27	184.204	164.0	27.0
CCCCCCCCCCCCCCCC	2.74	133.0708	456.0	10.0
CCCCCCCCCCCCCCCC				

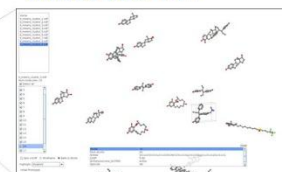
3D Visualization

Dataset Overview -- Clusters



- Datasets are separated into clusters, arranged in 3D space
- The intuitive interface of the 3D viewer allows to:
 - Zoom/rotate the clusters
 - Get valuable information on clusters via mouse over
 - Examine a cluster by clicking on it
 - The embedding into 3D space (positions/distance between clusters) reflects the similarity between clusters
 - Cluster can be removed from the dataset

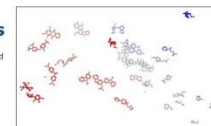
Inside Cluster View



- By selecting a cluster, the view zooms into the cluster and displays only the compounds included
- Details for each compound are available via mouse over
- Like the clusters, the compounds are embedded into the 3D space as well: the position/distance between compounds within the cluster reflects the similarity between compounds
- Compounds can be removed from the dataset

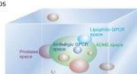
Highlight Features and Endpoints

- All compound properties can be highlighted
- The compounds are colored according to the numeric value, a high value is indicated by red, a low value is indicated by blue
- Also available for the cluster overview
- Gives an intuitive explanation towards the quality of the clustering approach: Does the clustering algorithm separate active from inactive compounds?



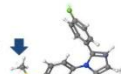
Cluster Compounds

- Compounds in the dataset are assigned to subgroups according to their similarity
- Supported cluster algorithms:
 - k-Means Clustering
 - Fixed number of k clusters
 - Random initialization, iterative update of clusters and cluster centroids
 - Hierarchical Clustering
 - Each compound is single cluster
 - Sequentially merge similar clusters
 - Structural Clustering
 - Find groups that share structural similarity
 - Compounds are assigned to clusters when there exists a common subgraph of sufficient size
- Developers can plug in new cluster algorithms



3D Alignment of Compounds

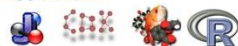
- Compounds in a cluster are likely to share common subgraphs:
 - This subgraph is already available if structural clustering is performed
 - Alternatively, the maximum common subgraph can be computed within each cluster
- The compounds within a cluster can be superimposed/aligned according to this subgraph:
 - This shows differences between compounds



Open-Source Webstart Application

- Java program that comes in two variants:
 - Java Web Start application (can directly started from a web browser)
 - Local installation that makes use of non-Java libraries
- Ches-Mapper is available at <http://opentox.informatik.uni-freiburg.de/ches-mapper>

Powered by:



References

- [1] Seifert, M., Griesch, T., Buchwald, F., Kramer, S. Online Structural Graph Clustering Using Frequent Subgraph Mining, 2010, Machine Learning and Knowledge Discovery in Databases, 213–228, Springer
- [2] Email an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org>

Acknowledgements

This work has been supported by the EU FP7 project 0HEALTH-F5-2008-200787 OpenTox (<http://www.opentox.org>).

Chemical Space Visualisation (Ches-Mapper)



CheS-Mapper: Chemical Space Mapping and Visualization in 3D



Abstract

Scientific researchers in the field of cheminformatics, chemical datasets. Therefore, the need for visualization. Our recently developed 3D molecular viewer Ches-Maps of 3D structural clustering, and multi-dimensional or similar compounds and consequently arranged in 3D similarity. This intuitively provides essential information allowing easy and understandable access to a large number of different clustering approaches employed in our descriptors of the compounds. These features can be understood the underlying scientific knowledge. As a first part of a given dataset for further analysis.

Build 3D Structure and Extract Features

- Select input dataset
 - Various dataset formats are supported (csv/molecules...)
 - Dataset can be directly loaded from the web
- 3D structure is built
 - 3D structure can be built with Chemical Development Kit (CDK) or OpenBabel
 - External libraries like Corina can be plugged in easily
- Extract features
 - Features are required for clustering and embedding
 - Automatic extraction of dozens of descriptors with CDK

Cluster Compounds

- Compounds in the dataset are assigned to subgroups according to their similarity
- Supported cluster algorithms:
 - k-Means Clustering
 - Fixed number of k clusters
 - Random initialization, iterative update of clusters and cluster centroids
 - Hierarchical Clustering
 - Each compound is single cluster
 - Sequentially merge similar clusters
 - Structural Clustering
 - Finds groups that share structural similarity
 - Compounds are assigned to clusters where there exists a common subgraph of sufficient size
- Developers can plug in new cluster algorithms

Home
k_means_cluster_1.sdf
k_means_cluster_2.sdf
k_means_cluster_3.sdf
k_means_cluster_4.sdf
k_means_cluster_5.sdf
k_means_cluster_6.sdf
k_means_cluster_7.sdf
k_means_cluster_8.sdf
k_means_cluster_9.sdf

☐ Spin on/off ☐ Wireframe ☒ Balls & Sticks

Highlight:

Initial Prototype

Cluster	k_means_cluster_9.sdf
Num molecules	19
LOGP	[2,81; 9,09] 0:5,09
ActivityOutcome_NCTRE	[active(19)]
Species	[rat(19)]
ActivityScore_NCTRE	[80; 100] 0:88,11

Chemical Space Visualisation (Ches-Mapper)



CheS-Mapper: Chemical Space Mapping and Visualization in 3D



Abstract

Scientific researchers in the field of chemoinformatics chemical datasets. Therefore, the need for visual: Our recently developed 3D molecular viewer Chet of 3D-atom structural clustering, and multi-dimes similar compounds and consequently arranged similarity. This intuitively provides essential info allowing easy and understandable access to a larg The different clustering approaches employed in descriptors of the compounds. These features ca understand the underlying scientific knowledge, a part of a given dataset for further analysis.

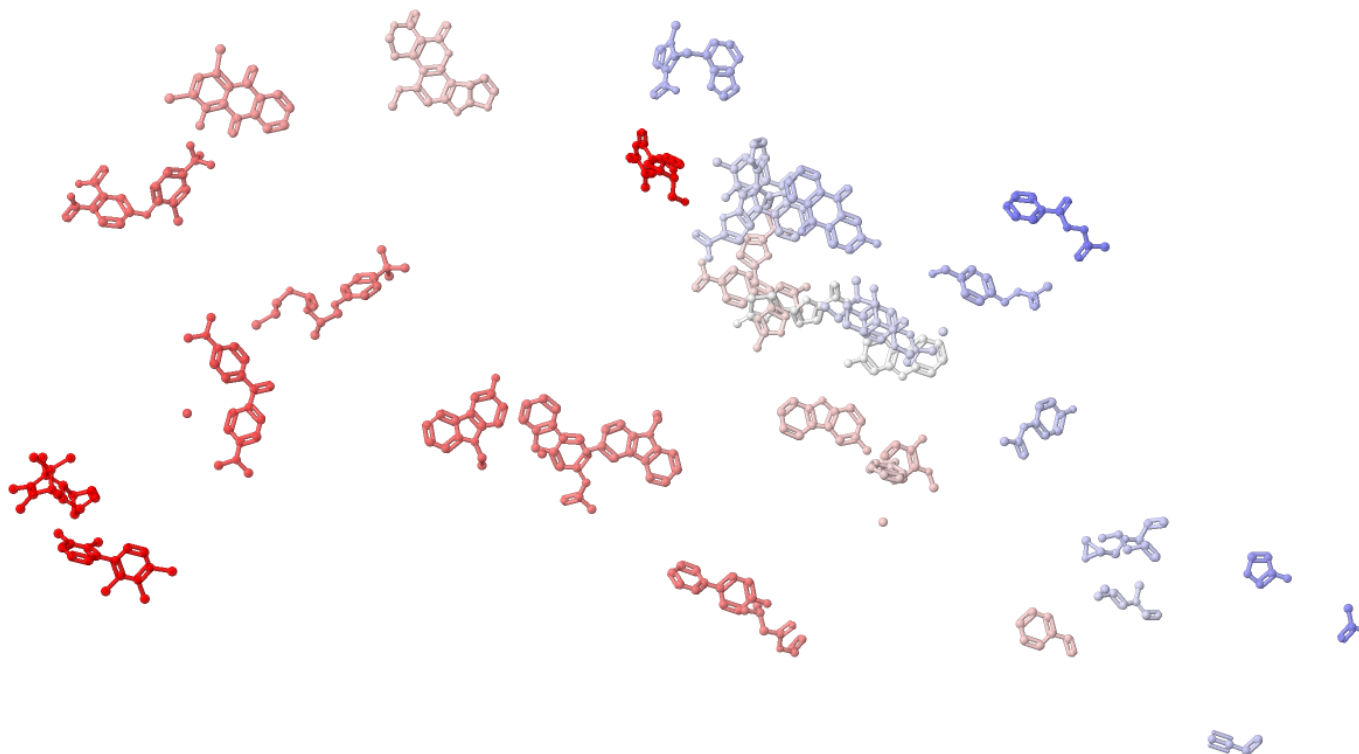
Build 3D Structure Extract Features

- Select input dataset
 - Various dataset formats are supported (csv/xml/indirect...)
 - Dataset can be directly loaded from the web
- 3D structure is built
 - 3D structure can be built with Chemical Development Kit (CDK) or OpenBabel
 - External libraries like Corina can be plugged in easily
- Extract features
 - Features are required for clustering and embedding
 - Automatic extraction of dozens of descriptors with CDK

Cluster Compound

- Compounds in the dataset are assigned to their similarity
- Supported cluster algorithms:
 - k-Means Clustering
 - Fixed number of k clusters
 - Random initialization, iterative up clusters and cluster centroids
 - Hierarchical Clustering
 - Each compound is single cluster
 - Sequentially merge similar cluster
 - Structural Clustering
 - Finds groups that share structural
 - Compounds are assigned to cluster there exists a common subgraph
- Developers can plug in new cluster algorithm

- Java program that comes in two variants:
 - Java Web Start application (can direct
 - Local installation that makes use of no
- Ches-Mapper is available at <http://open.tox>



ToxBank Infrastructure Project (started Jan 2011)

www.toxbank.net

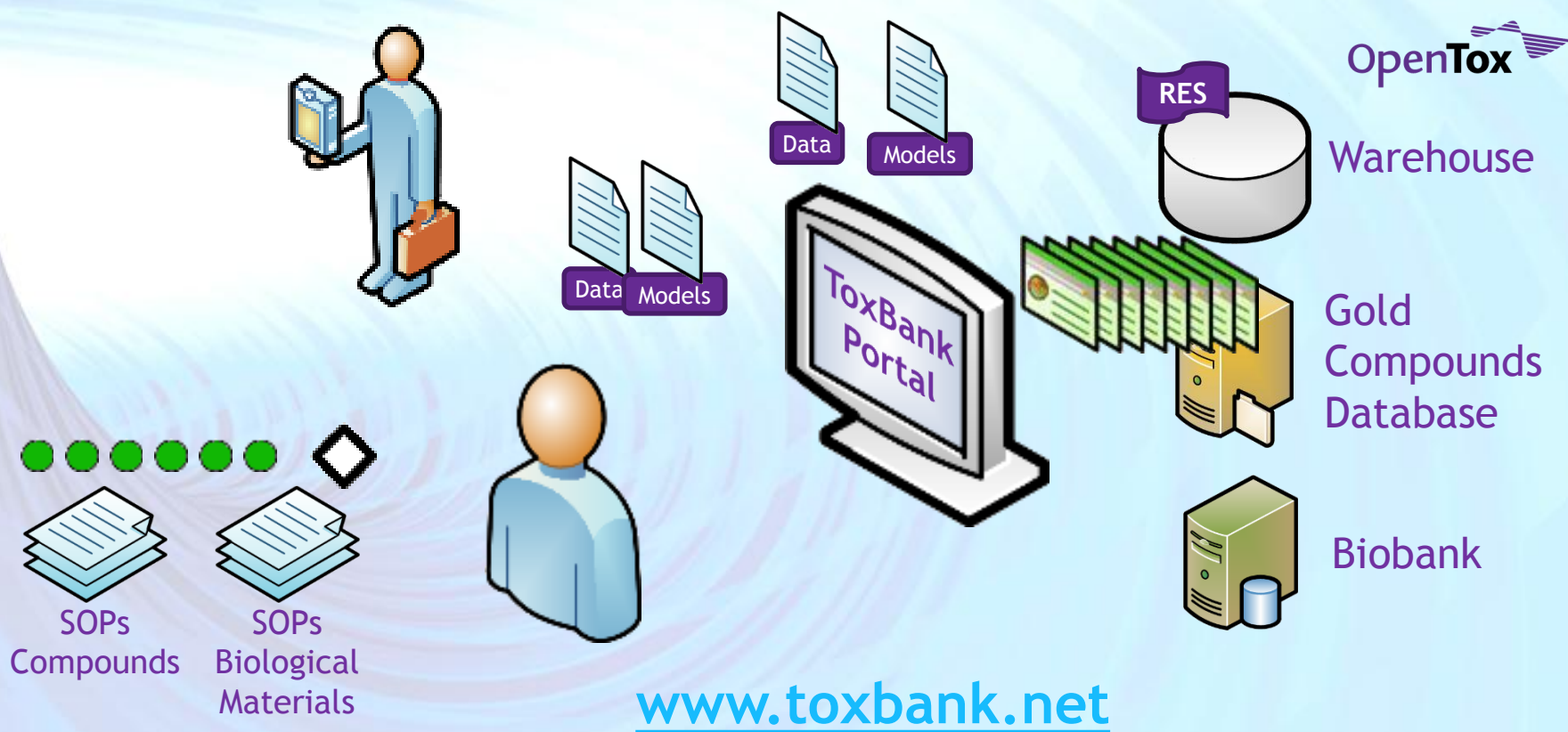
Establishment of a ...



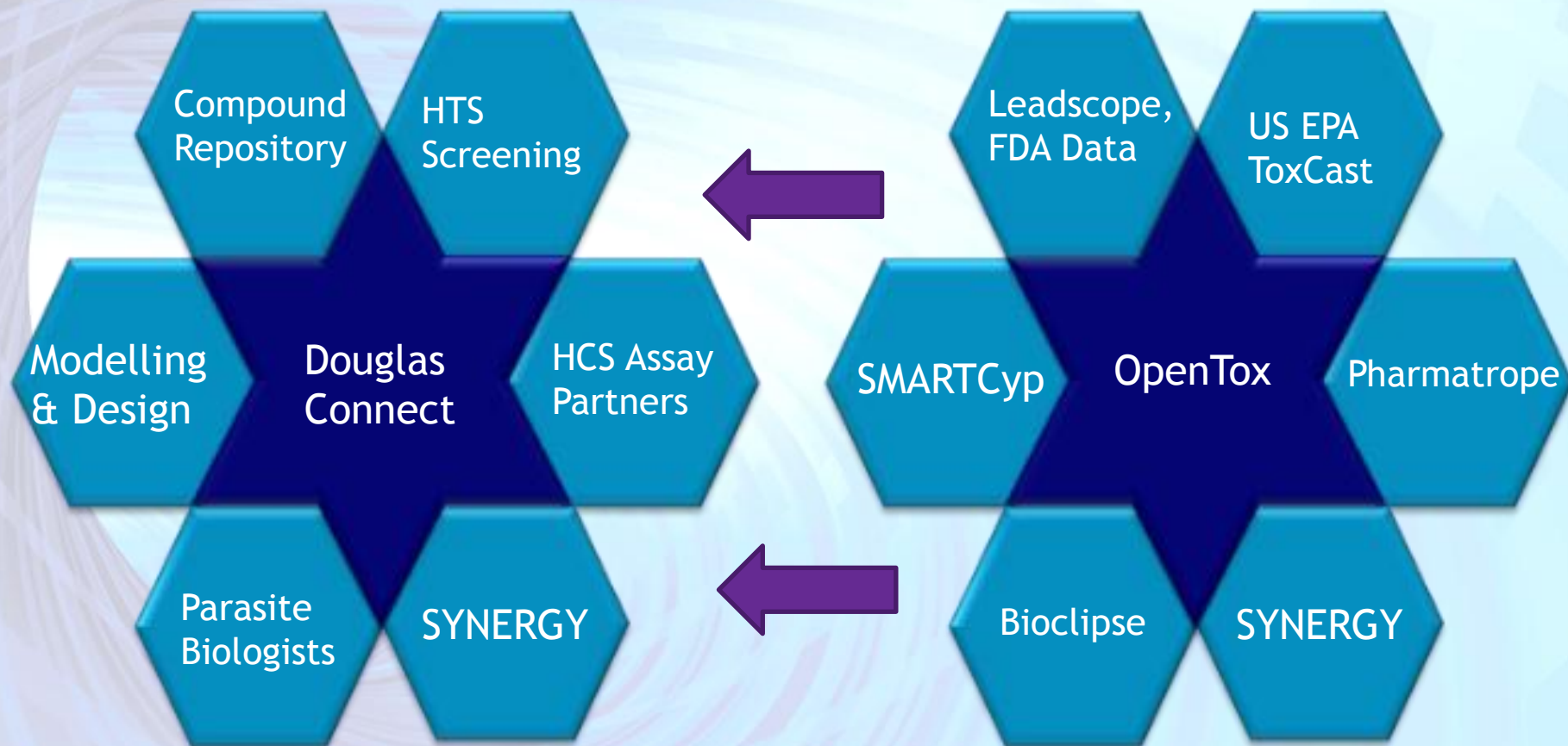
- ... cell and tissue banking information resource
- ... repository for the selected test compounds
- ... database of reference test compounds
- ... dedicated web-based data warehouse

Our Infrastructure Vision for ToxBank supporting all steps of Predictive Toxicology Research based on Alternative Testing methods

Users access compounds, biological materials, data and models for experimental planning and integrated analysis of experimental results



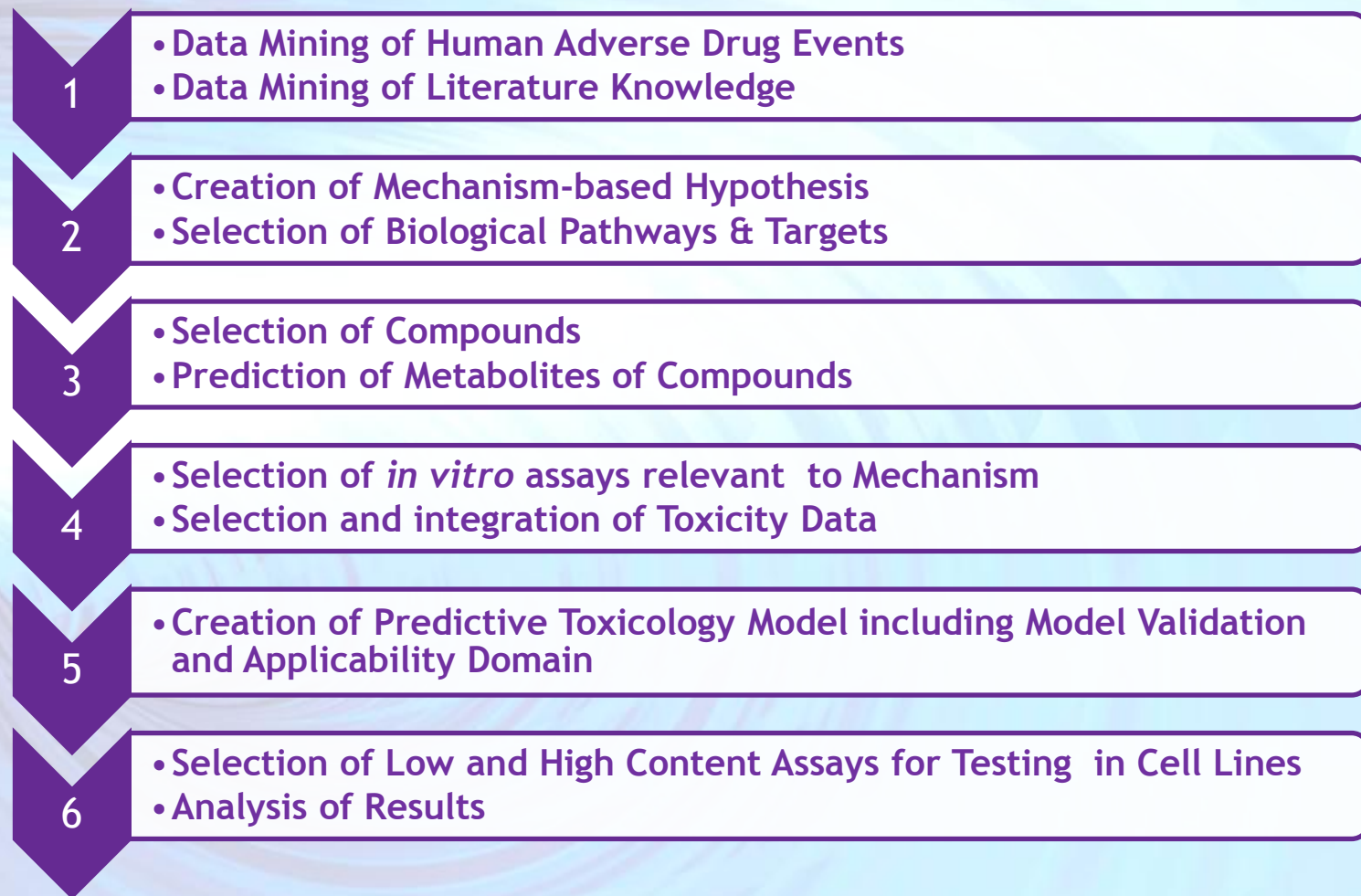
Virtual Organisation Pilots



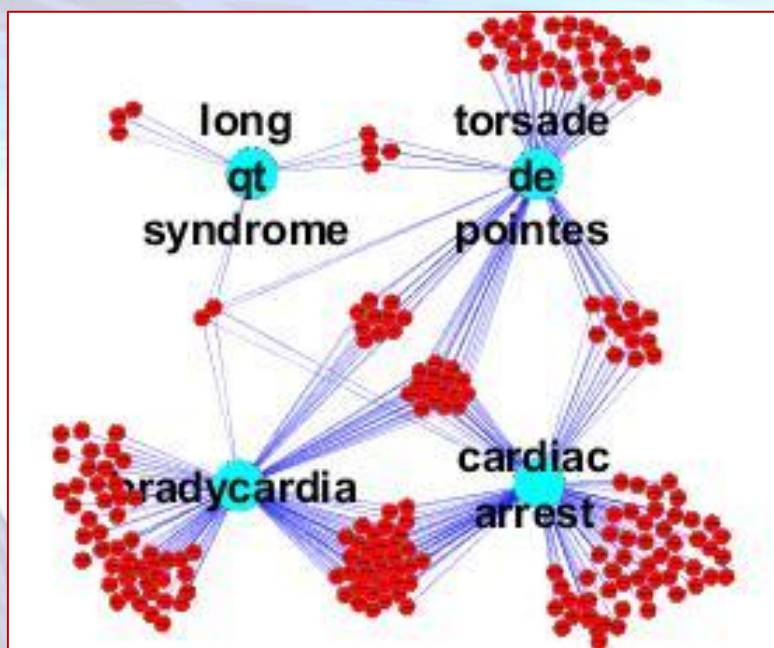
Neglected Disease Drug Design VO

Predictive Toxicology VO

OpenTox - Synergy Predictive Toxicology VO Pilot Strategy Development & Case Study



Analysis of Adverse Events Based on Pharmacological Activity



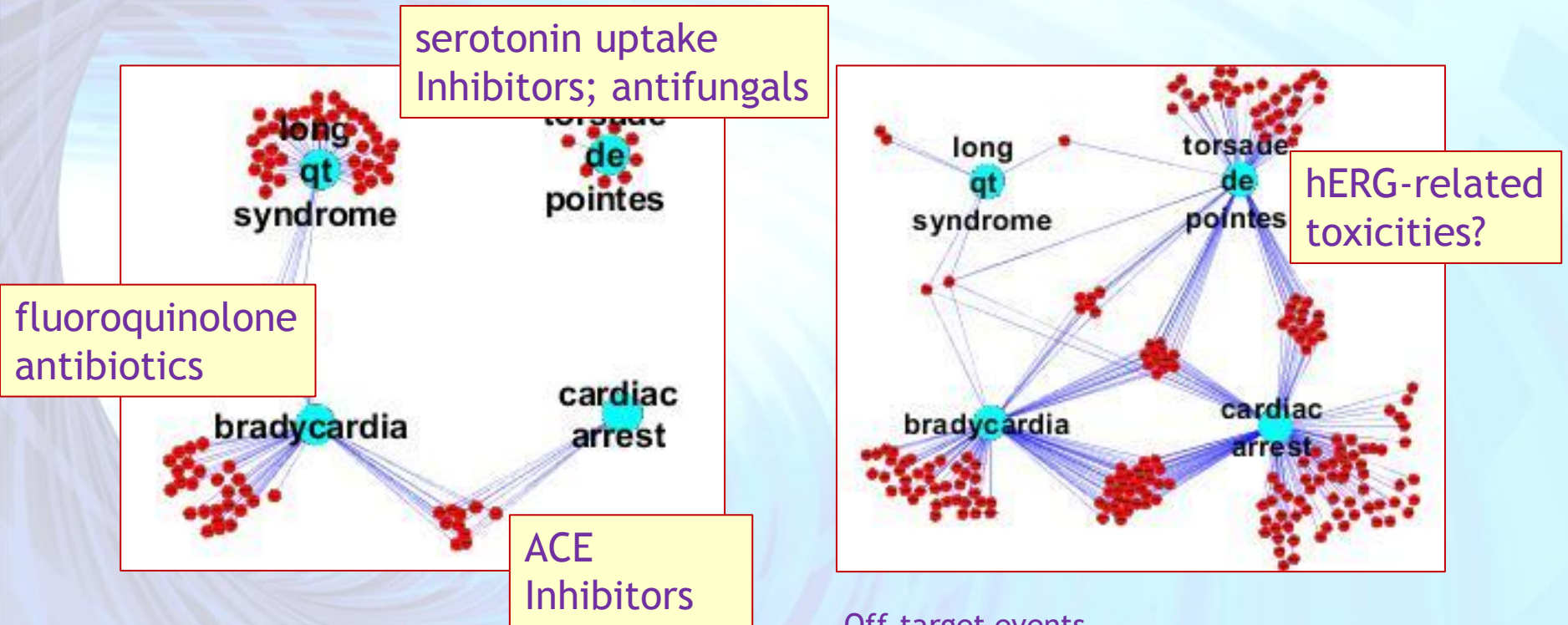
- Cardiac adverse events
- Related to hERG ion channel?

cyan = adverse event, red = drug
lines define links

- Question addressed:
 - Are the adverse events a function of inhibiting the pharmacological target?
 - Or is the adverse event due to an off-target activity?



Example: Cardiac Adverse Events



On-target events

cyan = adverse event, red = drug
lines define links

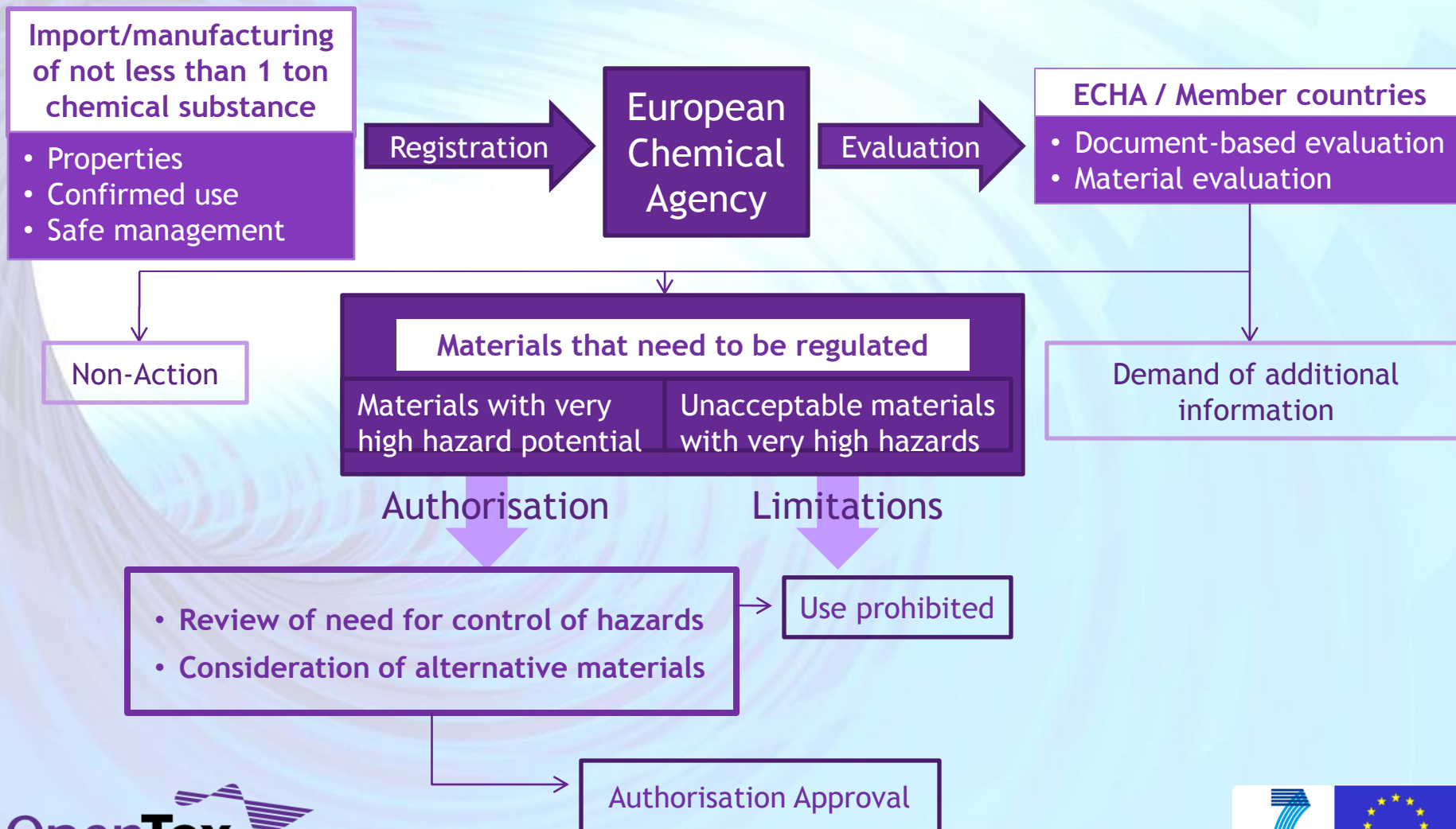
Off-target events



REACH



Introduction - REACH registration



REACH

“Registration, Evaluation, Authorisation & Restriction of Chemicals”

European Union Legislation, 2007

All chemicals imported or manufactured in Europe must be registered

from 1st December 2010 (European Chemicals Agency - ECHA)

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

Responsibility:

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

Threshold:

Phased:	2010	> 1,000 tons p.a.
	2018	> 1 ton p.a.



REACH



Registration per chemical: \$2M to \$14M²

IUCLID: International Uniform Chemical Information Database



Registration submitted by 22 Nov. 2010: 19,237¹
Chemicals pre-registered by 1 Dec. 2008: 143,000²



54M³ - 9M⁴ additional test animals

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

REACH

OpenTox InterAction Meeting Innovation in Predictive Toxicology

Modelling, Applications, REACH, Risk Assessment

9-12 August 2011

Technical University of Munich, Germany

Registration: Free but limited to 100 attendees

<https://www.surveymonkey.com/s/opentox2011>

Abstracts by: 30 April 2011, **Posters:** 30 June 2011

Bursary Award: three abstracts will be selected for travel & accommodation awards (by 30 April 2011)

More Information at:

www.opentox.org/meet/opentox2011



Collaborating Partners

In Silico Toxicology,
Switzerland

Douglas Connect,
Switzerland

Albert Ludwigs University
Freiburg, Germany

Ideaconsult,
Bulgaria

Istituto Superiore
di Sanità, Italy

Technical University
of Munich, Germany



National Technical
University of Athens,
Greece

Fraunhofer Institute
for Toxicology &
Experimental Medicine,
Germany

David Gallagher, UK

Institute of Biomedical
Chemistry of the Russian
Academy of Medical
Sciences, Russia

Seascope Learning &
JNU, India

Our Funding Support...

For more information, visit

www.opentox.org

Contact me:

barry.hardy -(at)- douglasconnect.com

twitter.com/barryhardy

barryhardy.blogs.com



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).