

Insights into An Interoperable Approach to Collaborative Drug Design and Predictive Toxicology

Barry Hardy (Douglas Connect)

Barry.Hardy@DouglasConnect.com

Presentation

24 February 2011

IQPC Predictive Toxicology
Conference, London, UK

Step 1: Search

Select structure(s)

Step 2: Verify structure

Verify structure

Step 3: Models

Select prediction models


Step 4: Estimate

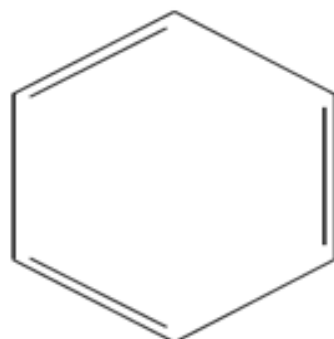
Estimate

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

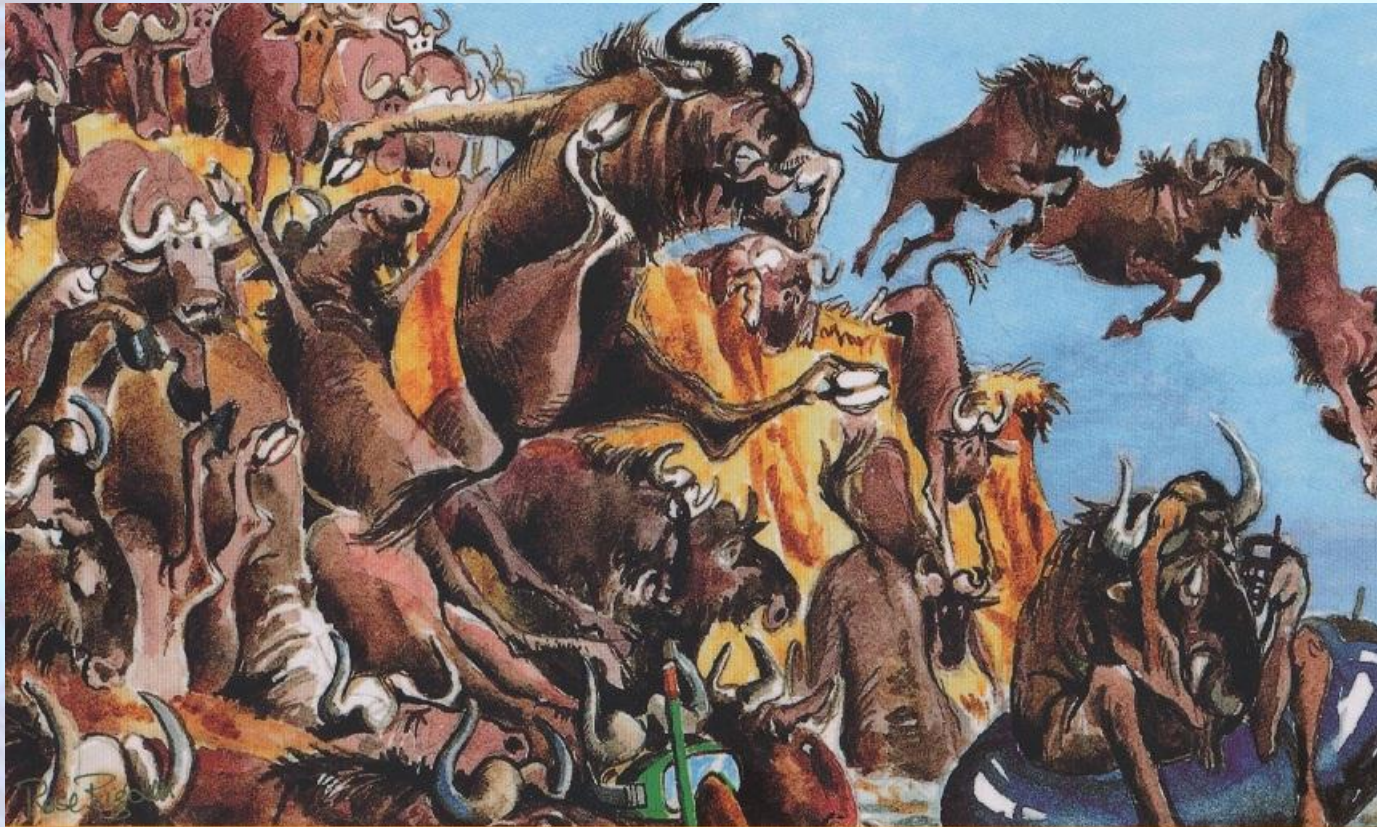
21742.0
 Benzene
 benzene
 OK

MolecularWeight  **MolecularWeight**

MW

78.1112

Migration

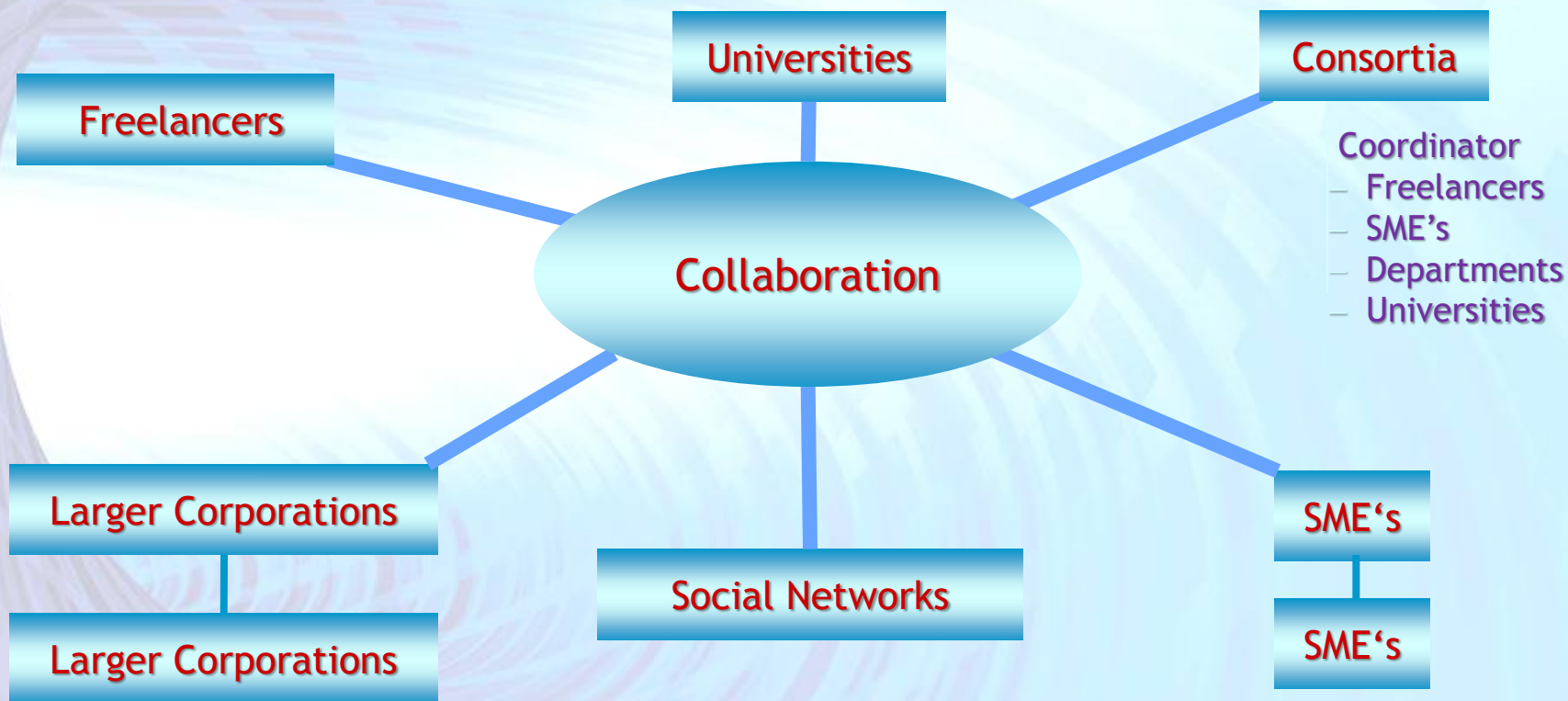


I'll get back to you, I'm in the middle of this migration thing!

Need for communications in the community overcoming different languages and vocabularies



Emerging drug discovery collaboration landscape

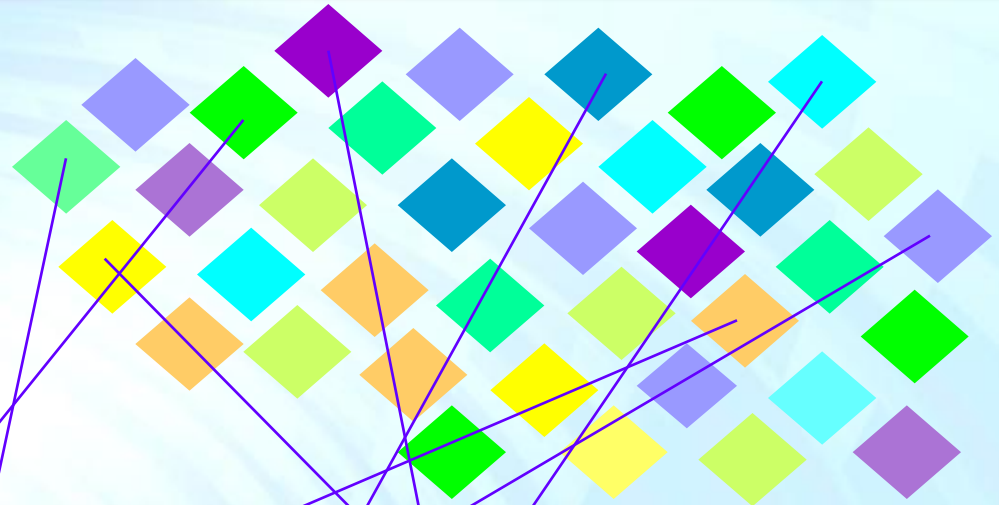


Creation of VO from Collaboration Pool

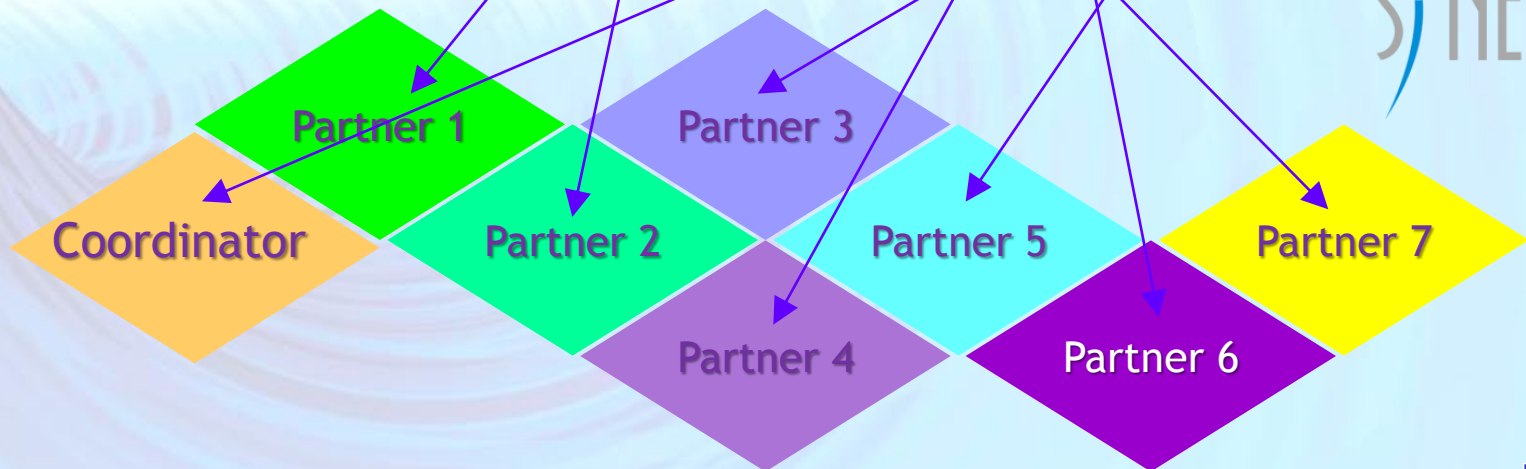
Network

Opportunity

Call for Tender
Need for joint effort
Major project

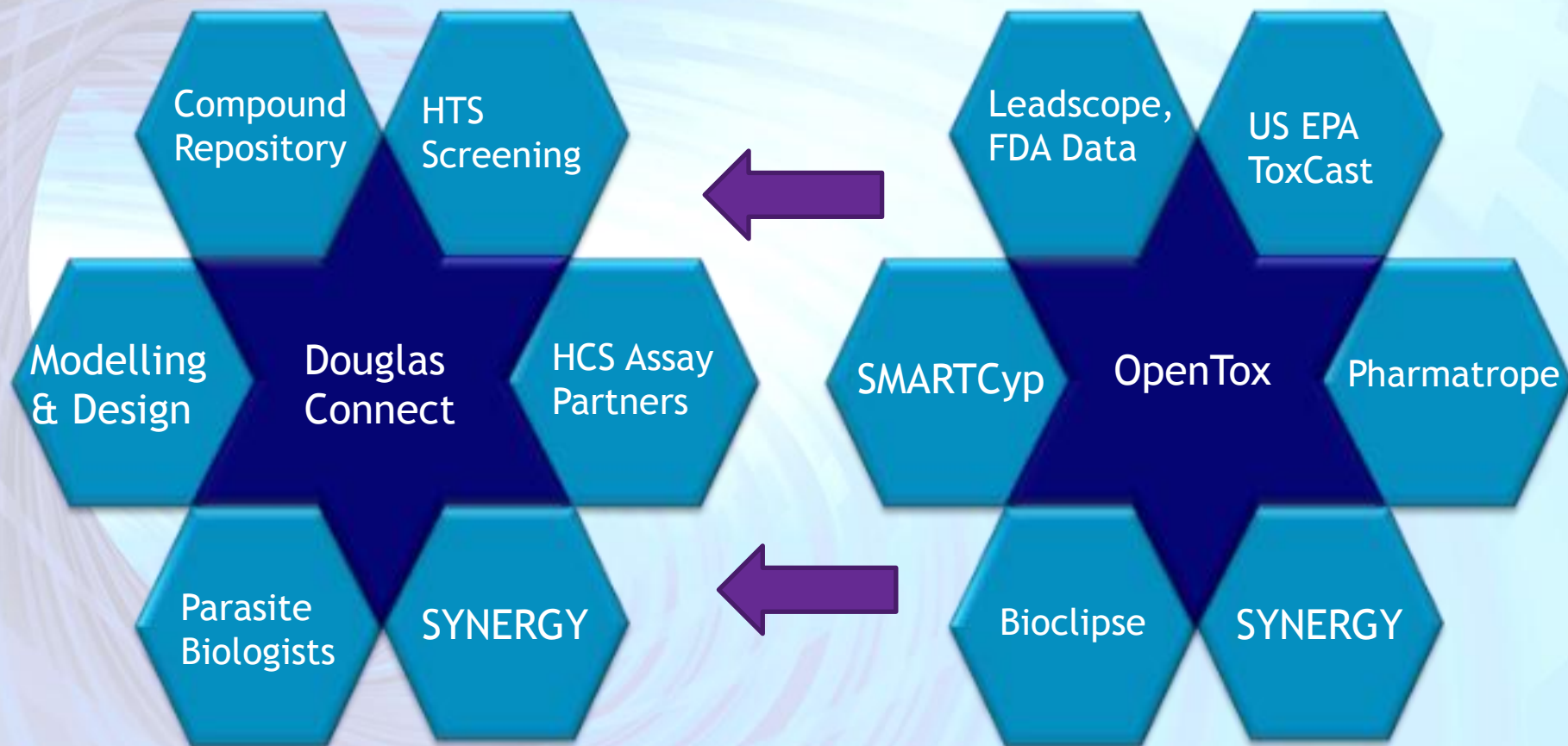


Virtual Organisation



SYNERGY

Virtual Organisation Pilots



Neglected Disease Drug Design VO

Predictive Toxicology VO

An African story



[SETAC Africa Conference \(31 May - 3 June\)](#)

Vision of our Founder Nicki Douglas - Applying our scientific developments to the benefit of all components of our environment



The “Tamboti Tree” Use Case



The “Tamboti Tree” Use Case



The “Tamboti Tree” Use Case



The “Tamboti Tree” Use Case



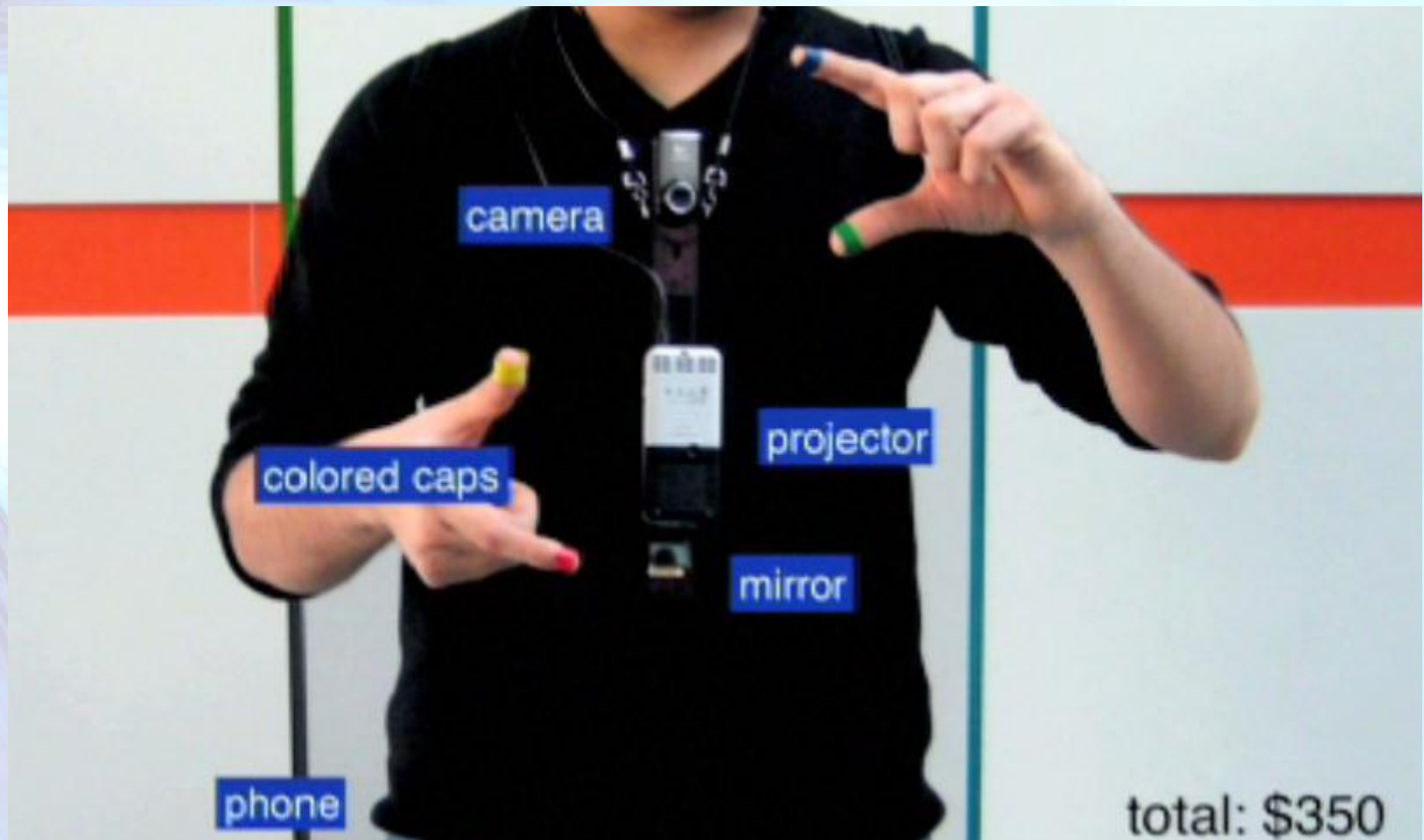
The “Tamboti Tree” Use Case



The “Tamboti Tree” Use case



Augmented Reality



Processing Packaging Information



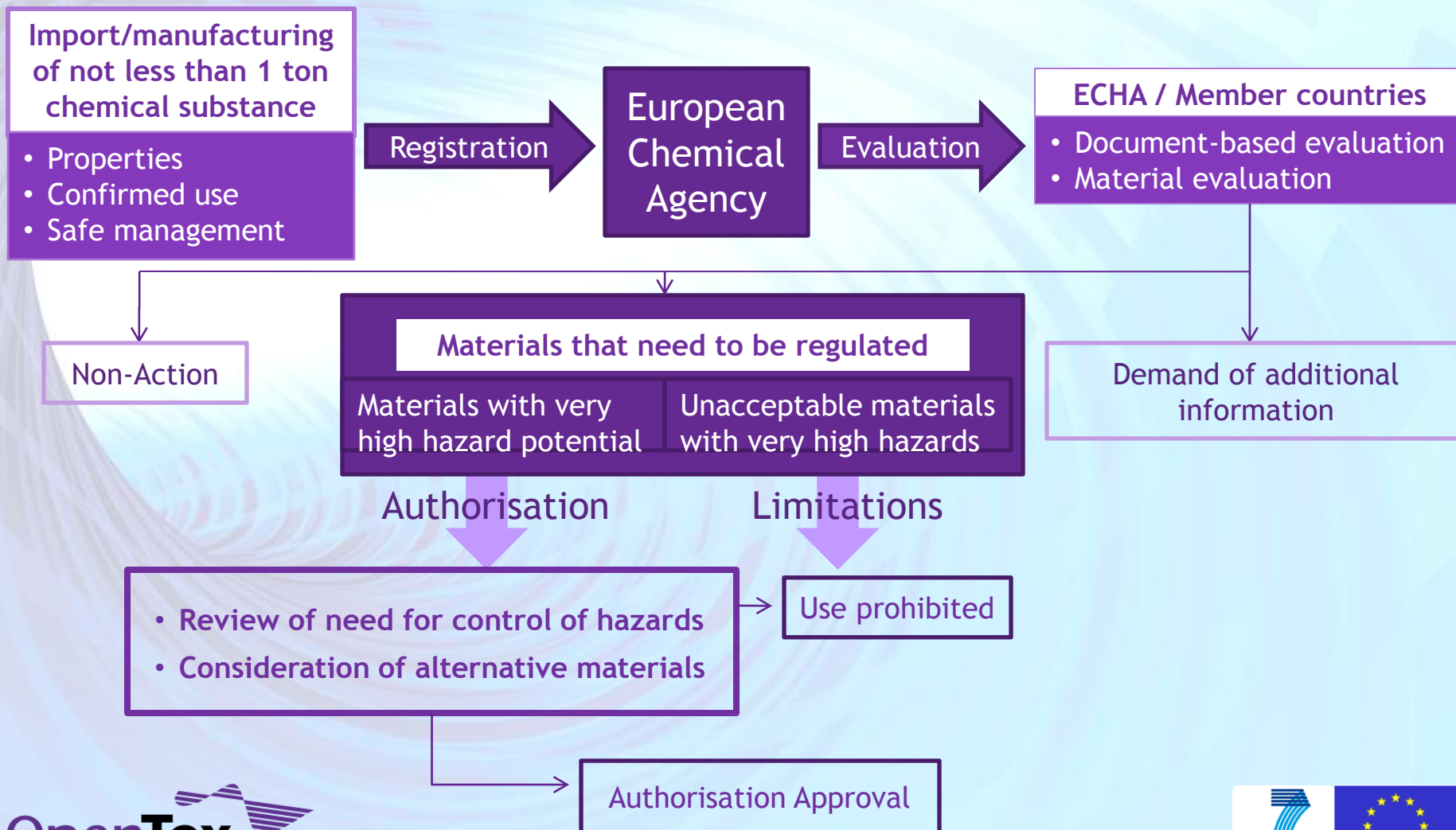
MIT Media Lab

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

Introduction - REACH

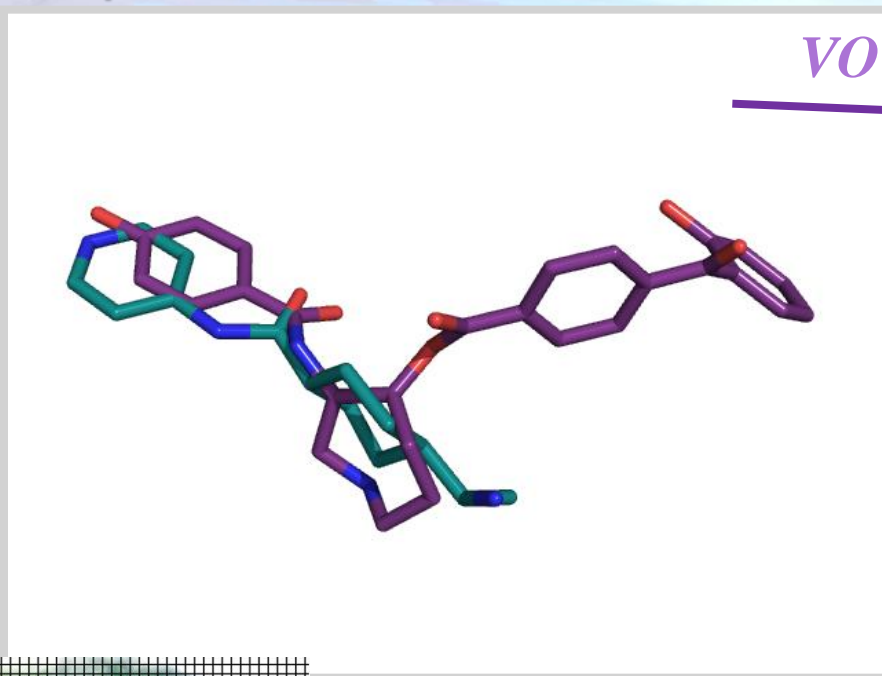


Introduction - REACH registration



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

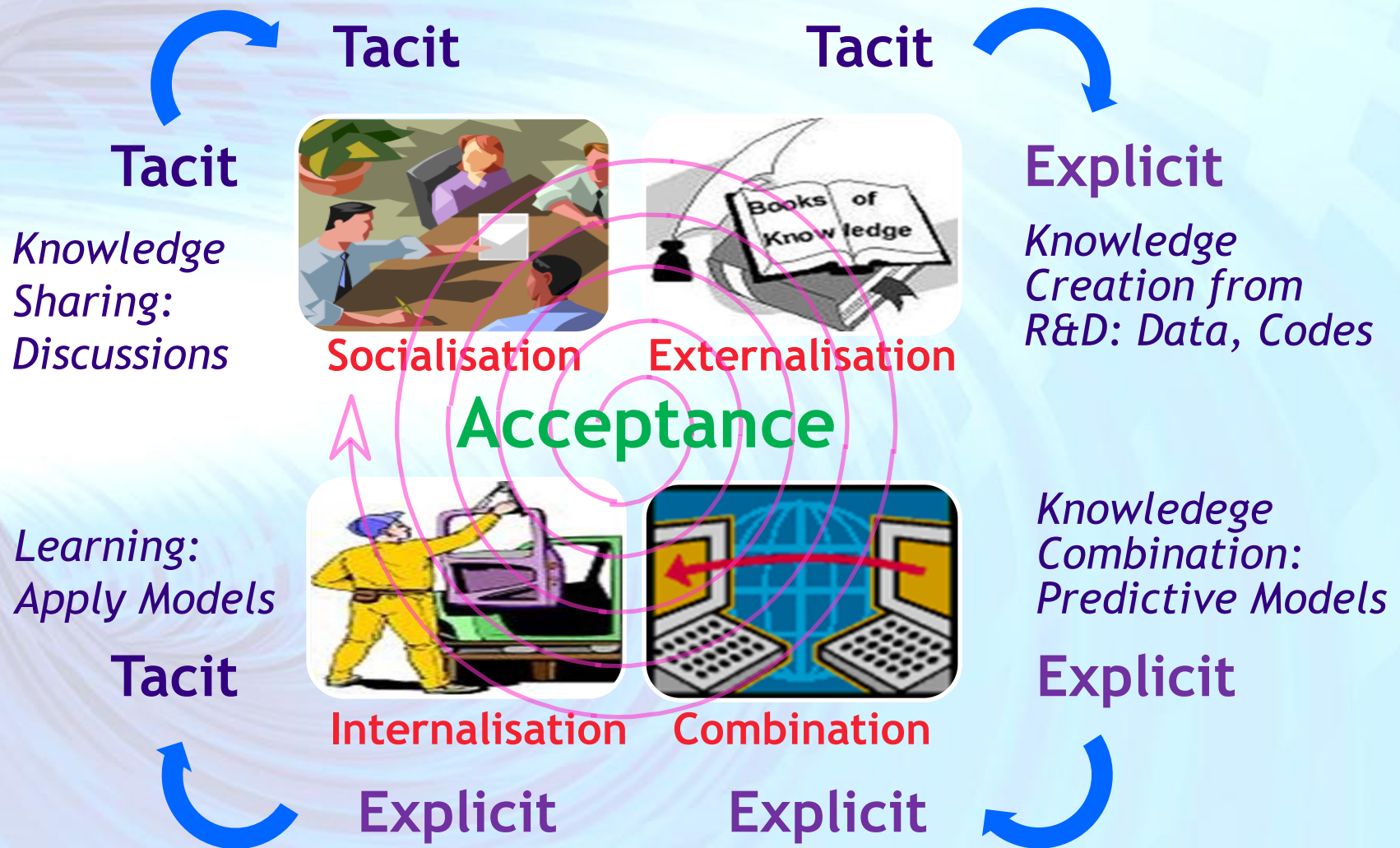
REACH Reporting
(QPRF, QMRF)

Applicability
Domain

Validation

Human
Data

SECI Model for Knowledge Management



Complexity Context

Non Repeatable
Adaptative, Patterns,
Filters

Sense
Making for
Emergent
Practice

Leadership
Novel
Practice

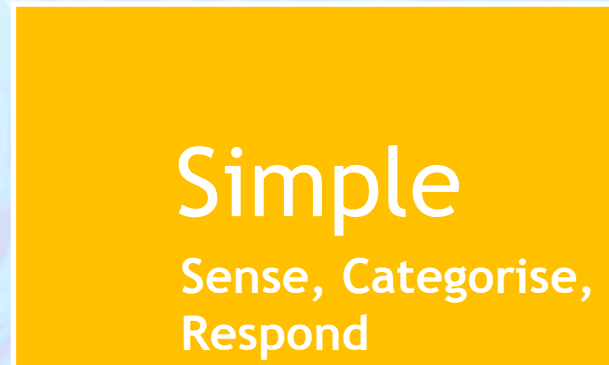
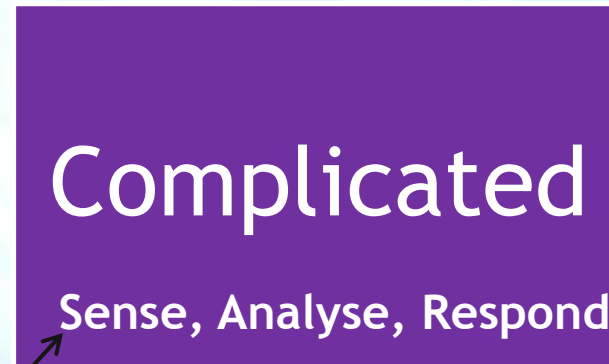
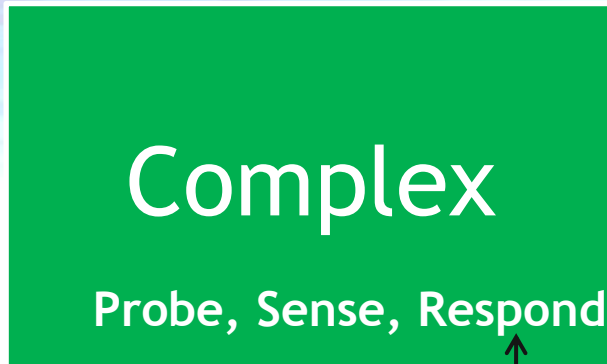
Lack of Cause & Effect, Stability-focused
Intervention, Crisis Management

Complex Cause & Effect
Systems Thinking, Analysis

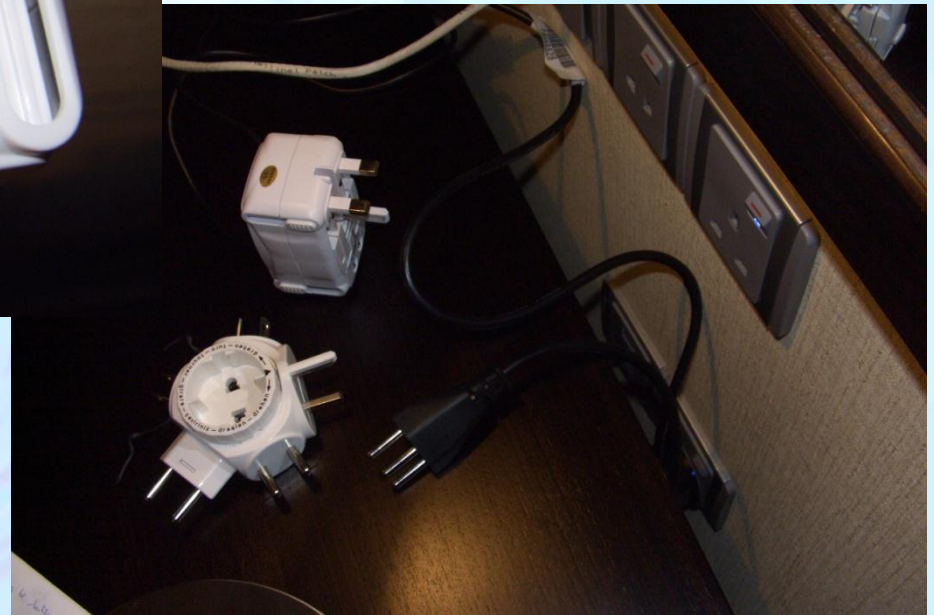
Processes
Good
Practice

Procedures
Best
Practice

Cause & Effect
Repeatable, SOPs



Ontology and Data - Interoperability

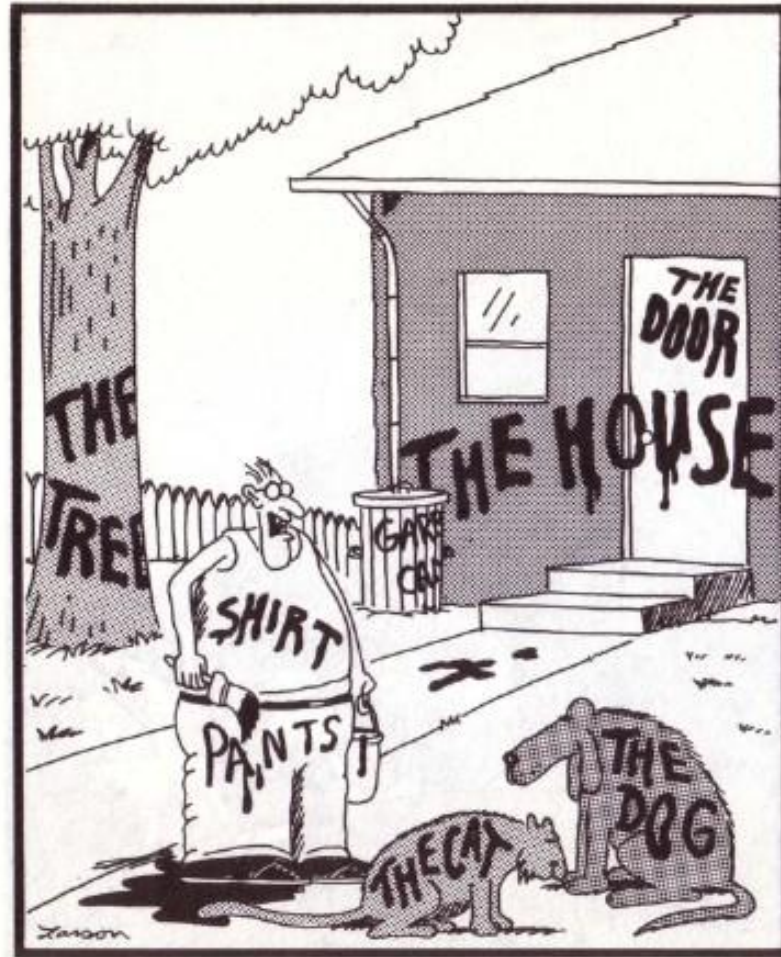


Interacting Components create Solutions



Adaptor Solution in Jeddah, 2008

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

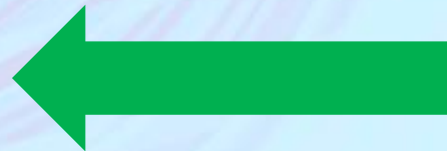


Siamese



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**)



Linked Data approach can also be applied to other resource types e.g., for algorithms or models as done in OpenTox... **Linked Resource** approach enables Knowledge Creation, Combination and Analysis

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

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

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

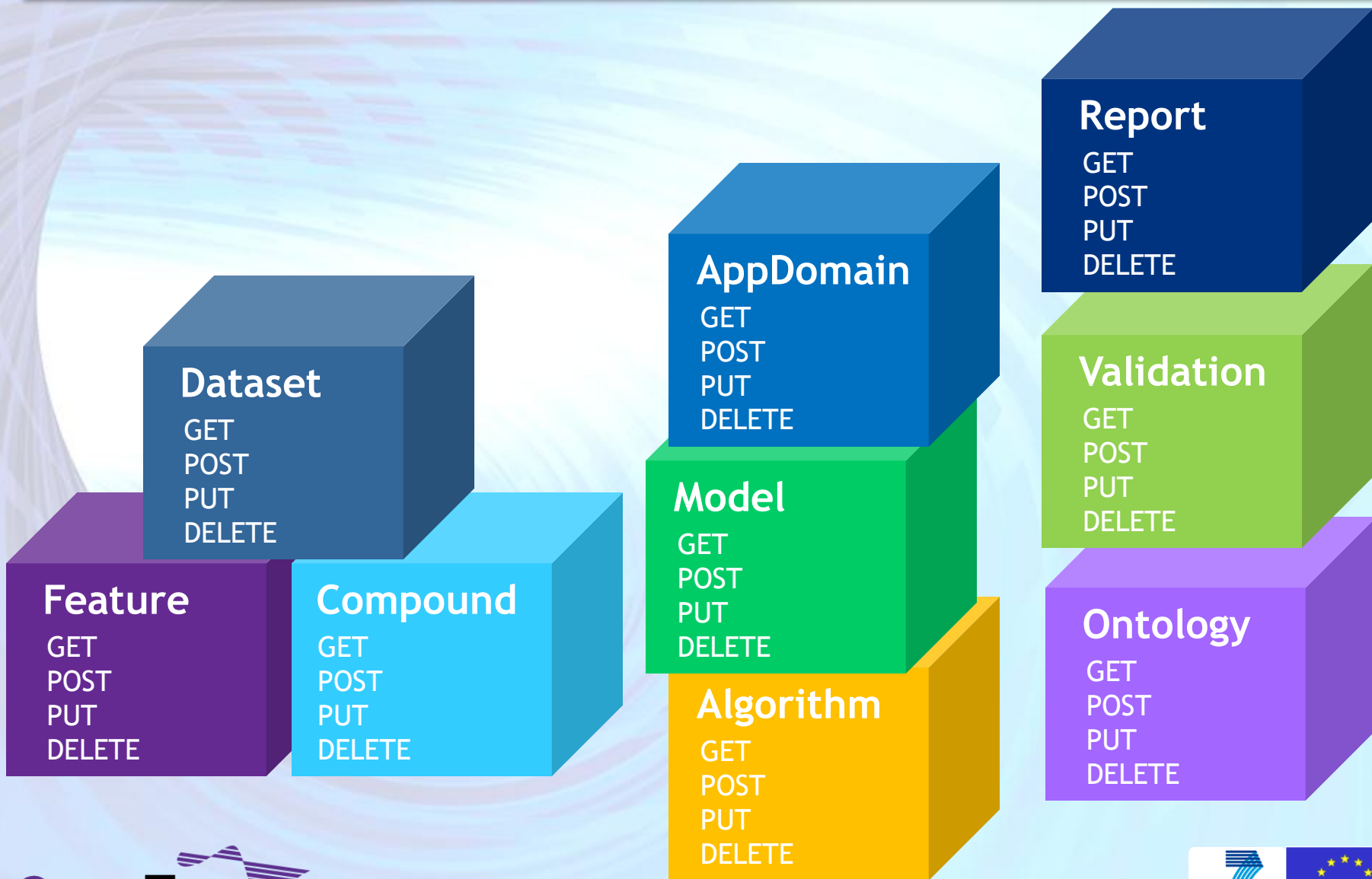
Validation: Methods for estimating the accuracy of model predictions

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

	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

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

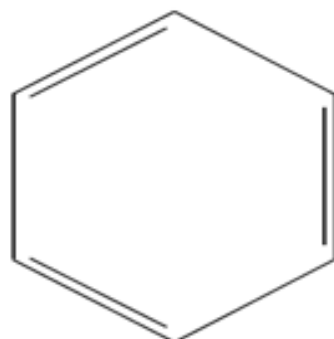
Step 3: Models
Select prediction models

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

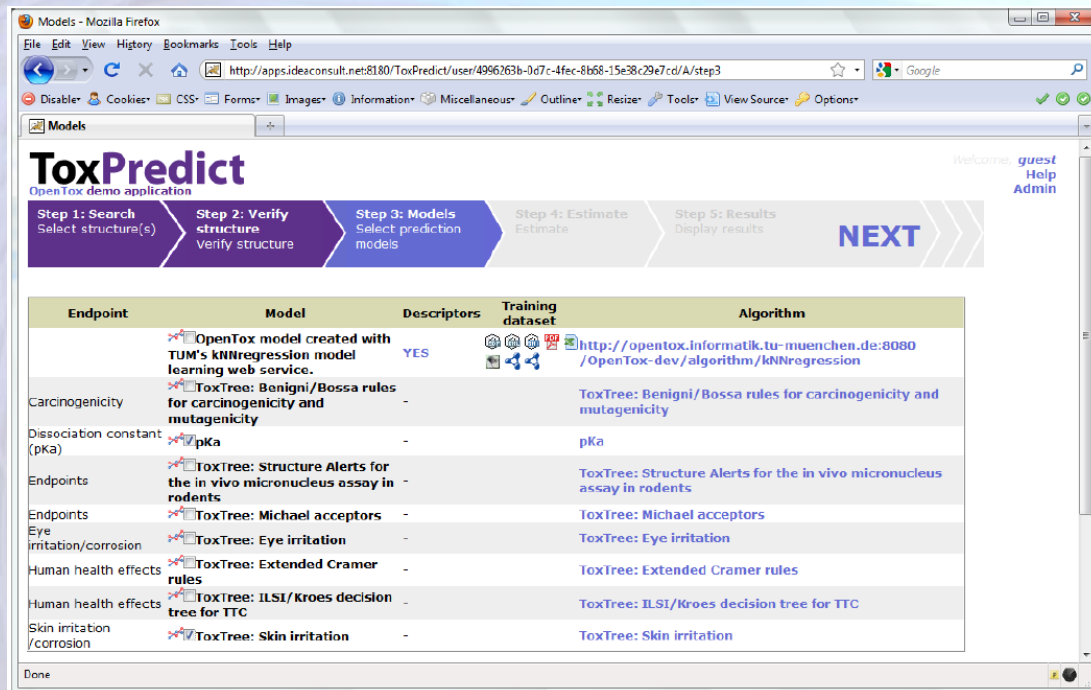
21742.0
Benzene
benzene
OK


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 has three steps: Step 1: Search (Select structure(s)), Step 2: Verify (Verify structure), and Step 3: Models (Select prediction models). Below the steps is a table with columns: Endpoint, Model, and Descriptors.

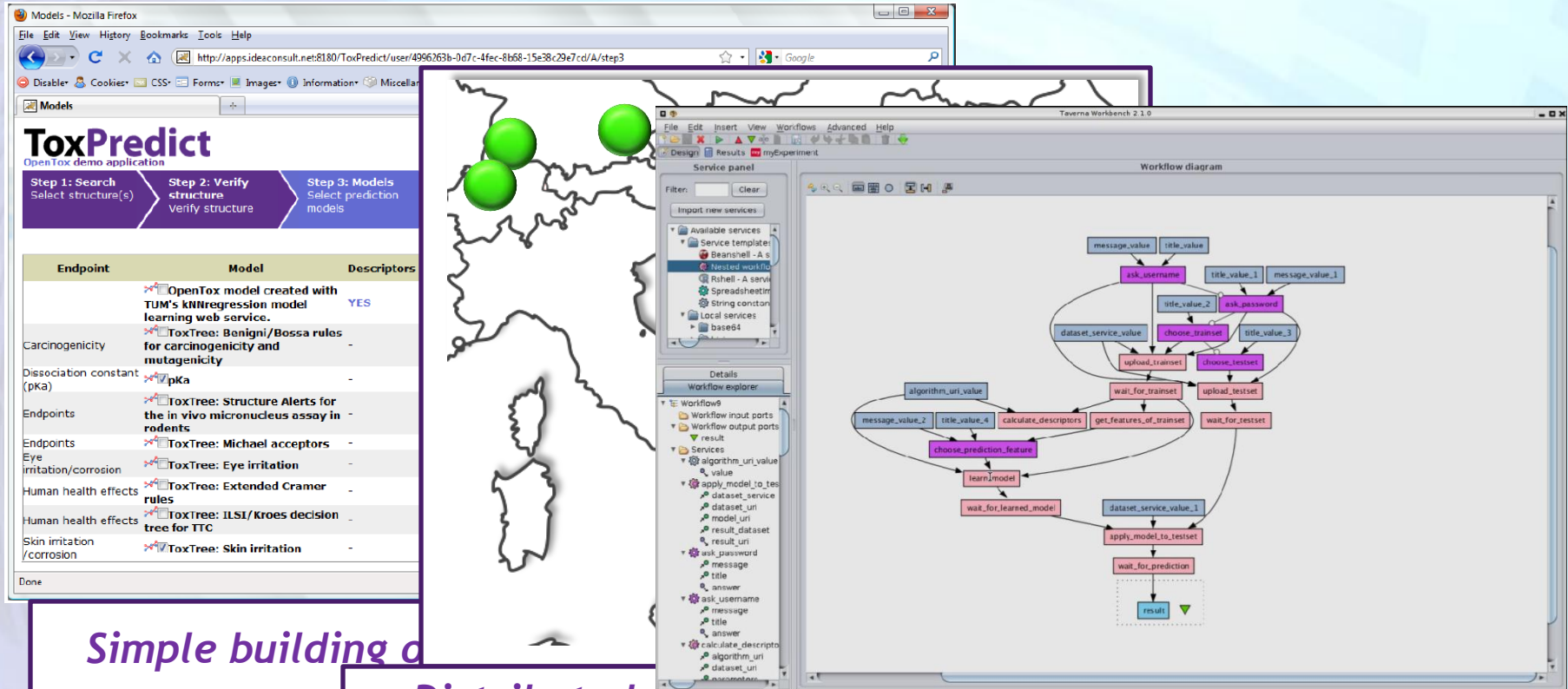
Endpoint	Model	Descriptors
	OpenTox model created with TUM's kNNregression model learning web service.	YES
Carcinogenicity	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	-
Dissociation constant (pKa)	pKa	-
Endpoints	ToxTree: Structure Alerts for the in vivo micronucleus assay in rodents	-
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 area), and two in Southern Europe (Spain/Italy area).

Simple building of applications methods and

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

What you can do with it ...



The image displays two overlapping windows. The background window is a Mozilla Firefox browser showing the ToxPredict web application. The foreground window is the Taverna Workflow editor showing a workflow diagram for model training and prediction.

ToxPredict Web Application:

OpenTox demo application

Step 1: Search
Select structure(s)

Step 2: Verify
structure
Verify structure

Step 3: Models
Select prediction
models

Endpoint	Model	Descriptors
Carcinogenicity	OpenTox model created with TUM's kNN regression model learning web service.	YES
Dissociation constant (pKa)	ToxTree: Benigni/Bossa rules for carcinogenicity and mutagenicity	-
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	-

Done

Taverna Workflow Diagram:

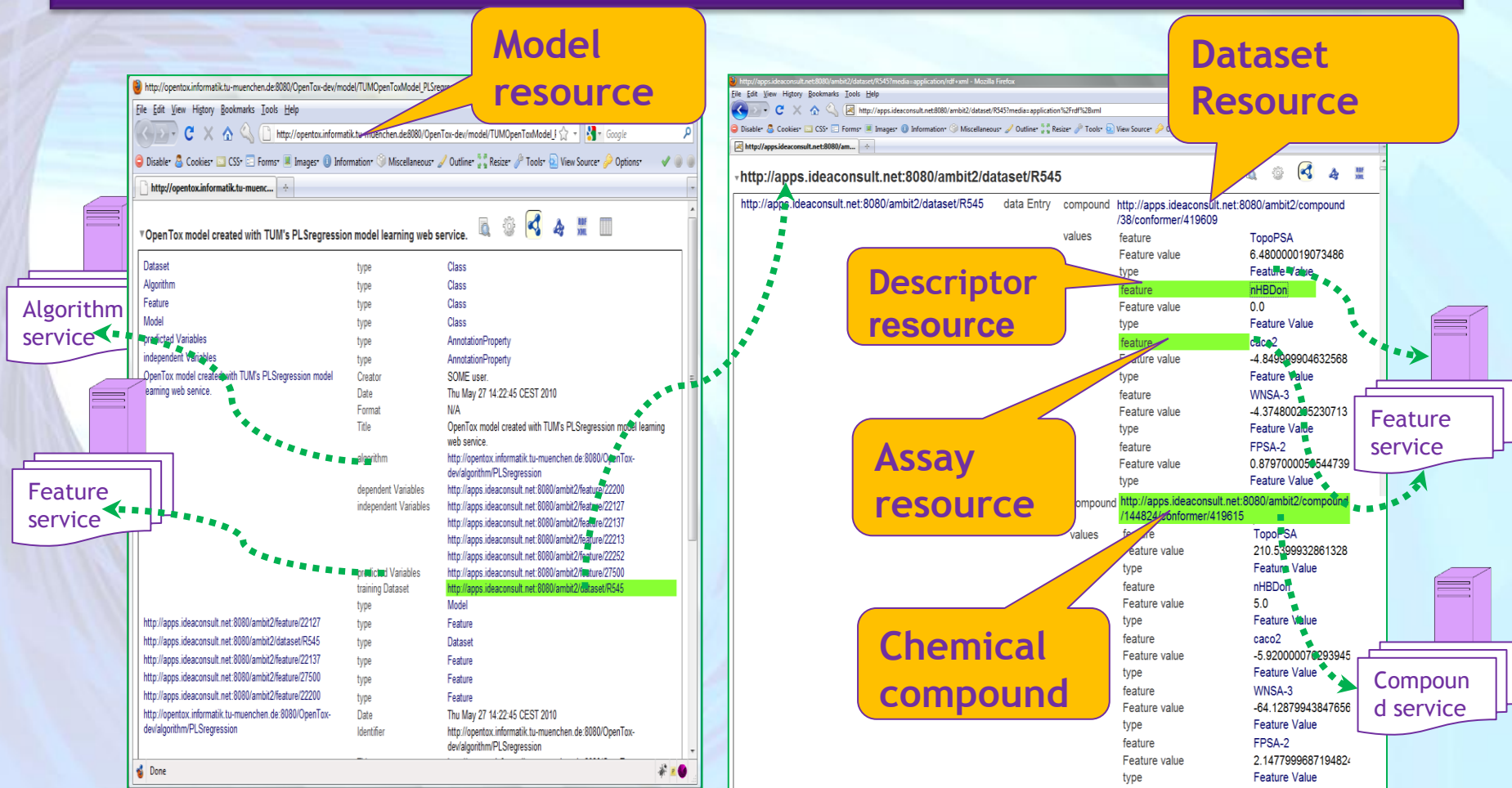
The workflow diagram illustrates a process for model training and prediction. It starts with input services (ask_username, title_value, message_value) leading to dataset_service_value. This is followed by upload_tramset, choose_tramset, and upload_testset. The process then moves to wait_for_tramset, get_features_of_tramset, and wait_for_reset. The next steps are calculate_descriptors, choose_prediction_feature, learn_model, and wait_for_learned_model. Finally, the workflow proceeds to dataset_service_value_1, apply_model_to_testset, wait_for_prediction, and ends with a result output.

Simple building of
applications
methods and

Distributed of
wide range of
methods

Integration into workflow systems for
computational biology

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



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

Dataset
Resource

Descriptor
resource

Assay
resource

Chemical
compound

Blue Obelisk
algorithms
ontology

Regression
Classification
Quantum
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	nHBDdon
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	FP5A-2
Feature value	0.8797000050544739
type	Feature Value
compound	http://apps.ideaconsult.net:8080/ambit2/compound/144824/conformer/419615
values	TopoPSA
Feature value	210.5399938861328
type	Feature Value
feature	nHBDdon
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	FP5A-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
nHBDdon	type	DatatypeProperty
Source	sameAs	http://www.blueobelisk.org/ontologies/chemoinformatics-algorithms/#nHBDdonors
Title		nHBDdon
Source		http://apps.ideaconsult.net:8080/ambit2/algorithm/org.openscience.cdk.qsar.descriptors.molecular.HBondDonorCountDescriptor
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	sameAs	caco2
Units		c049084m_caco2-training_set.sdf
type		Numeric Feature
sameAs		Gastrointestinal absorption

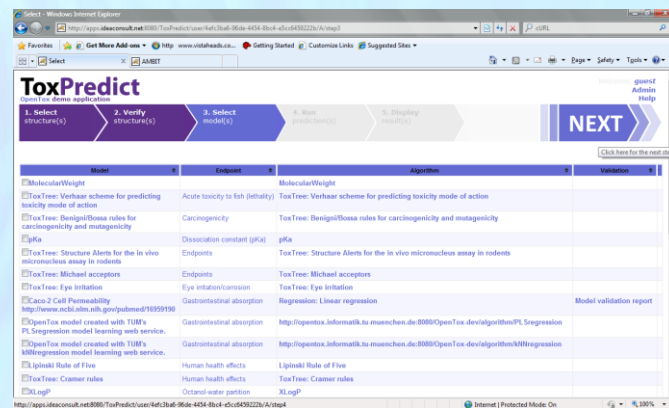
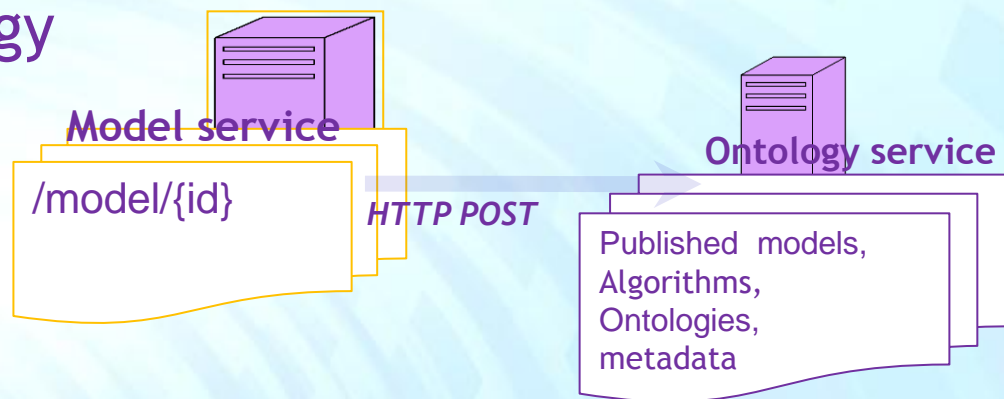
Make the model available

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

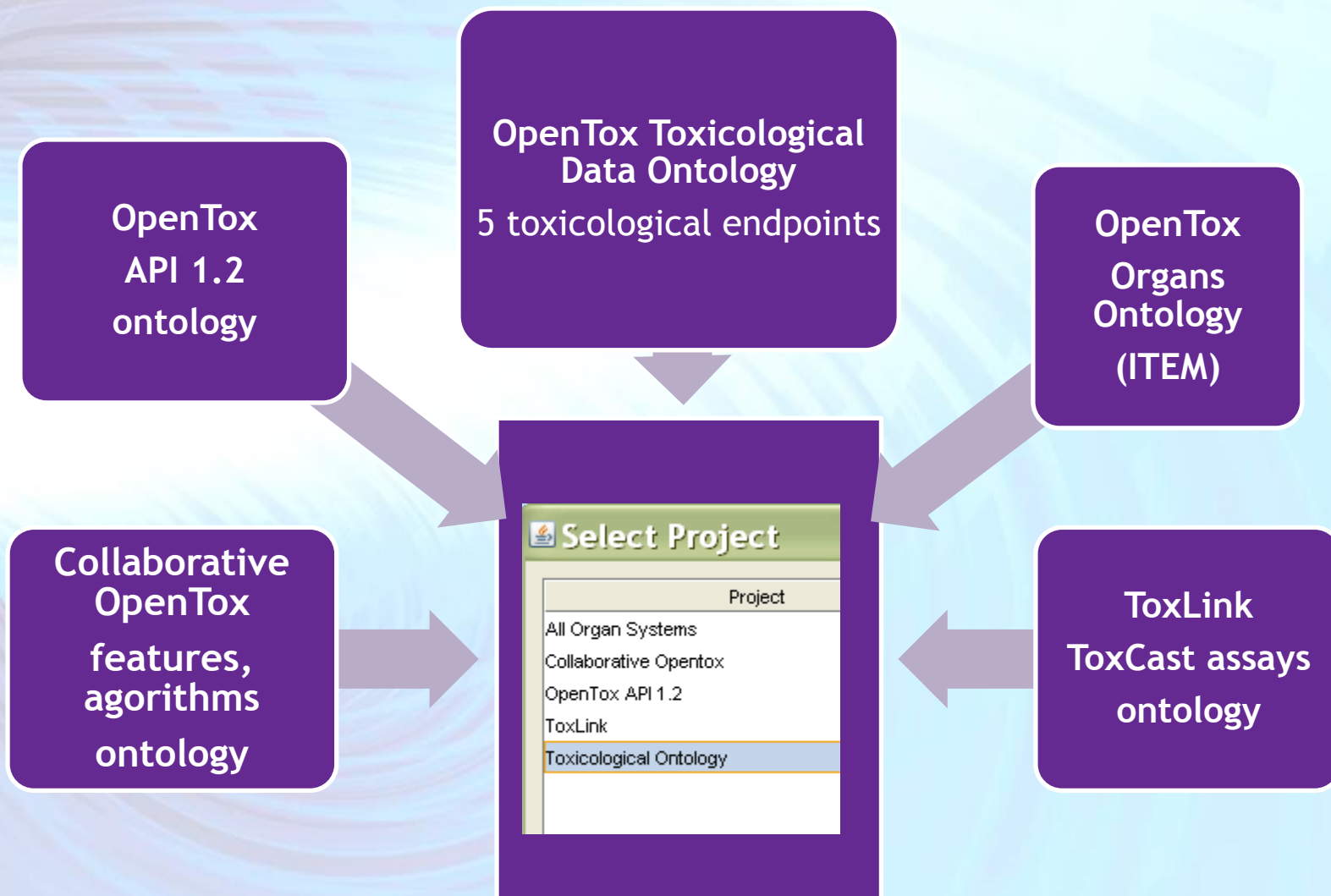


Need for communications in the community overcoming different languages and vocabularies



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

Collaborative Ontology Development: Collaborative Protege Server



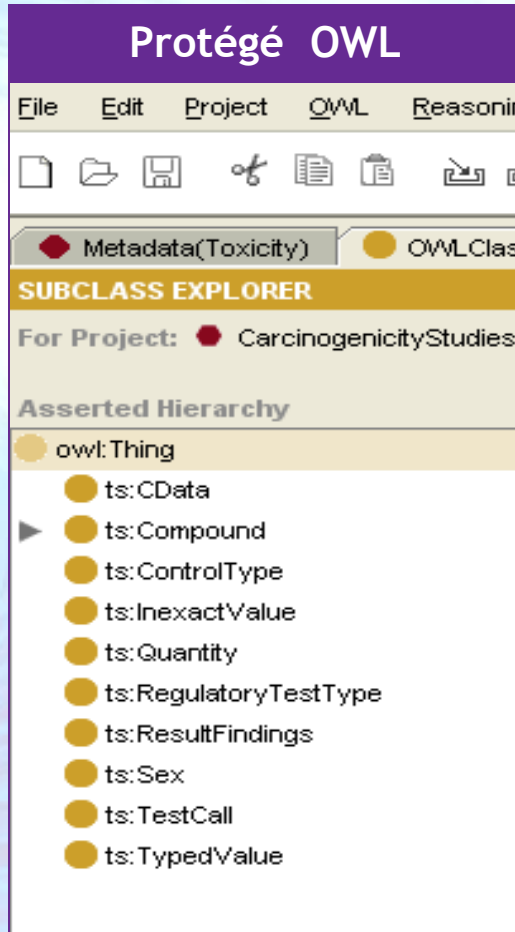
ToxML: conversion in OWL ontology

- 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

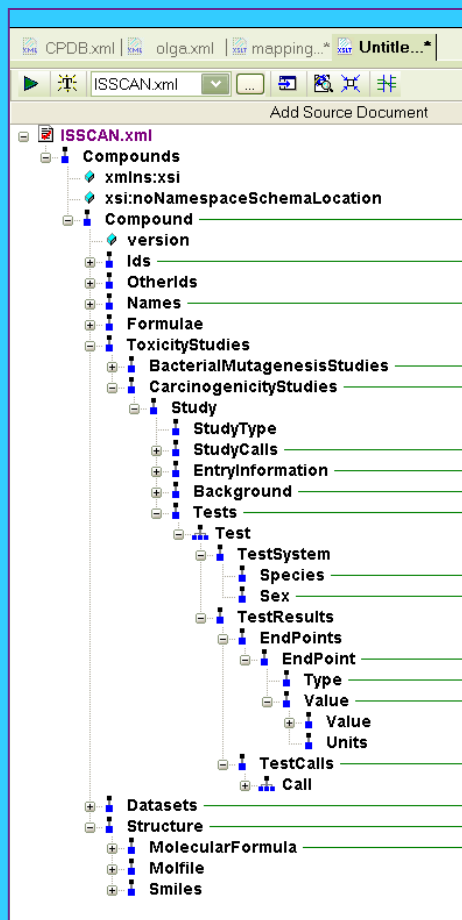
the parent relationships are kept

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

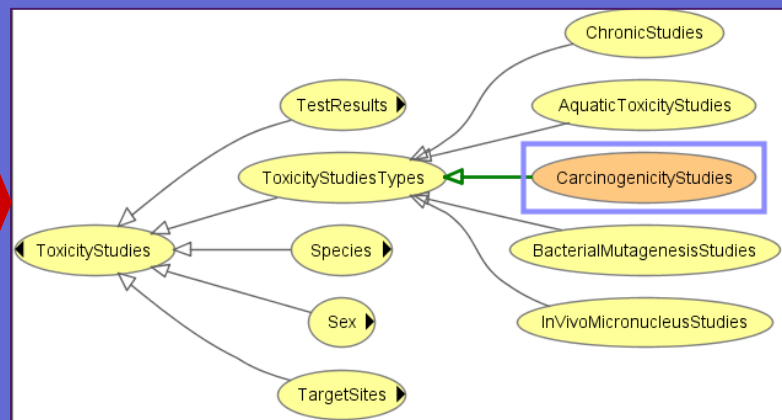
ToxML schema



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

OpenTox
Toxicological
Endpoint
Ontology

Ontology Development

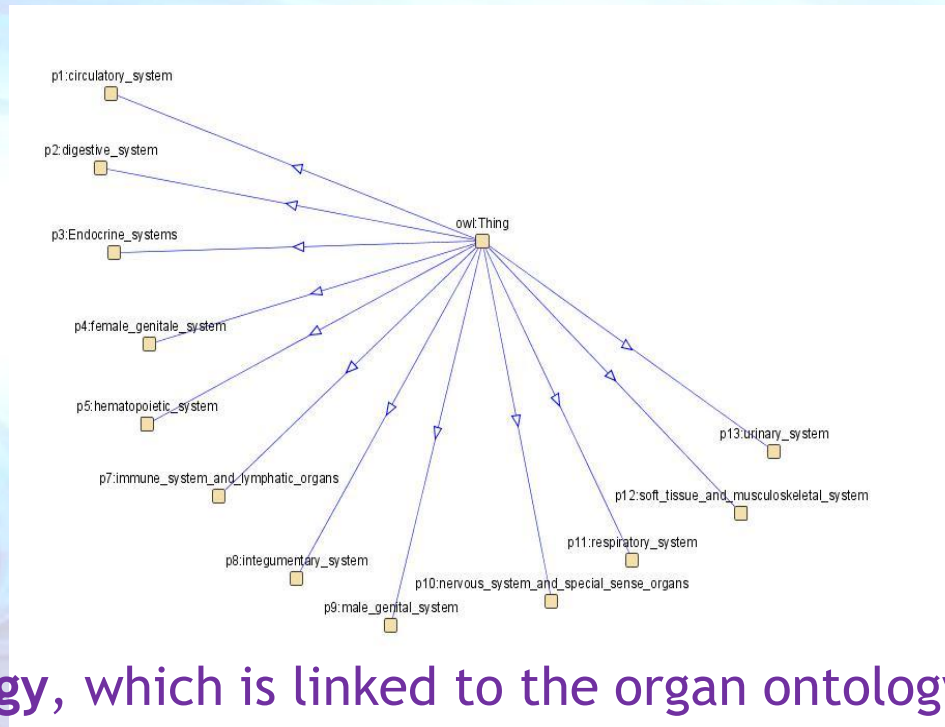


Re-use of terms defined in
neighbouring ontologies (e.g. OBO)

Collaborative
Protégé
Environment

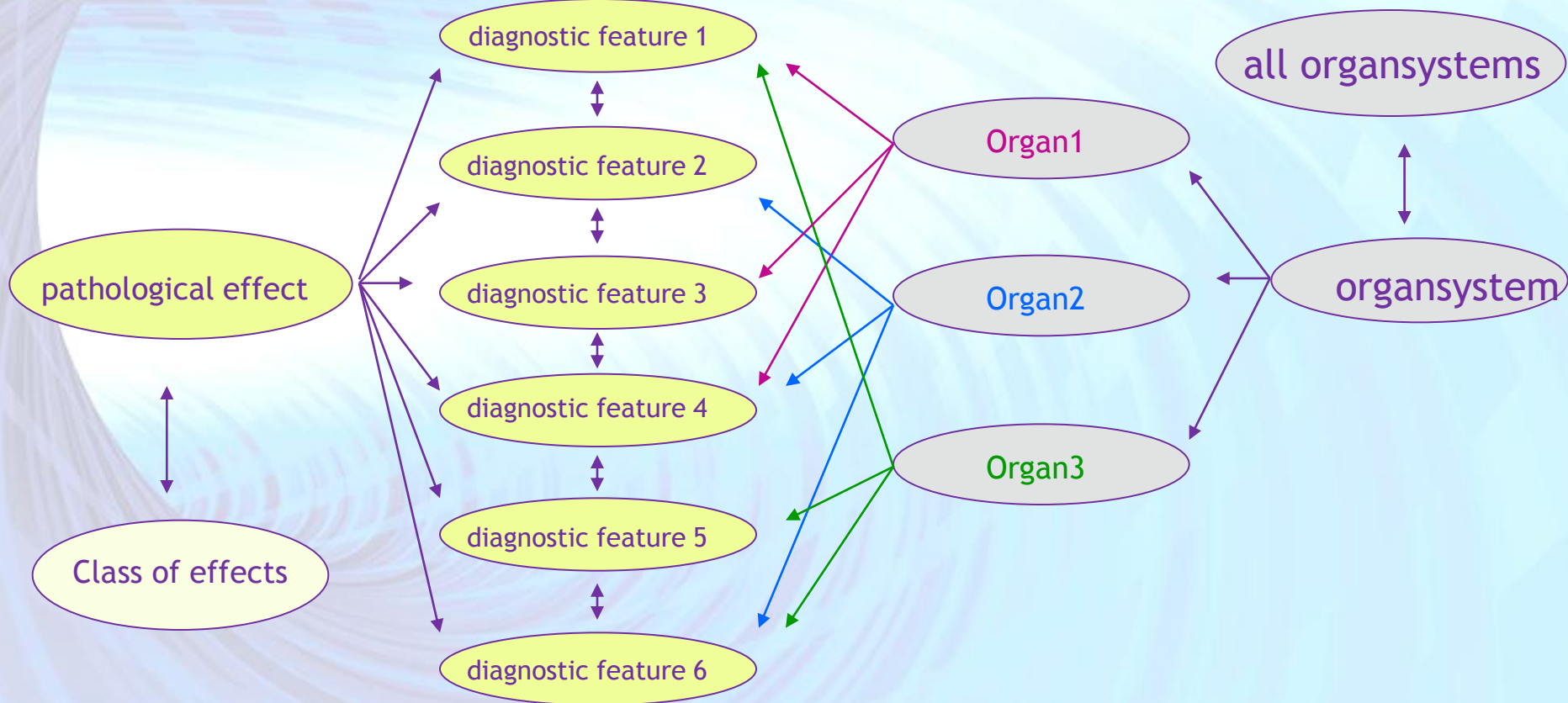
OpenTox Organ Ontology Development

- 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



Need for communications in the community overcoming different languages and vocabularies



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

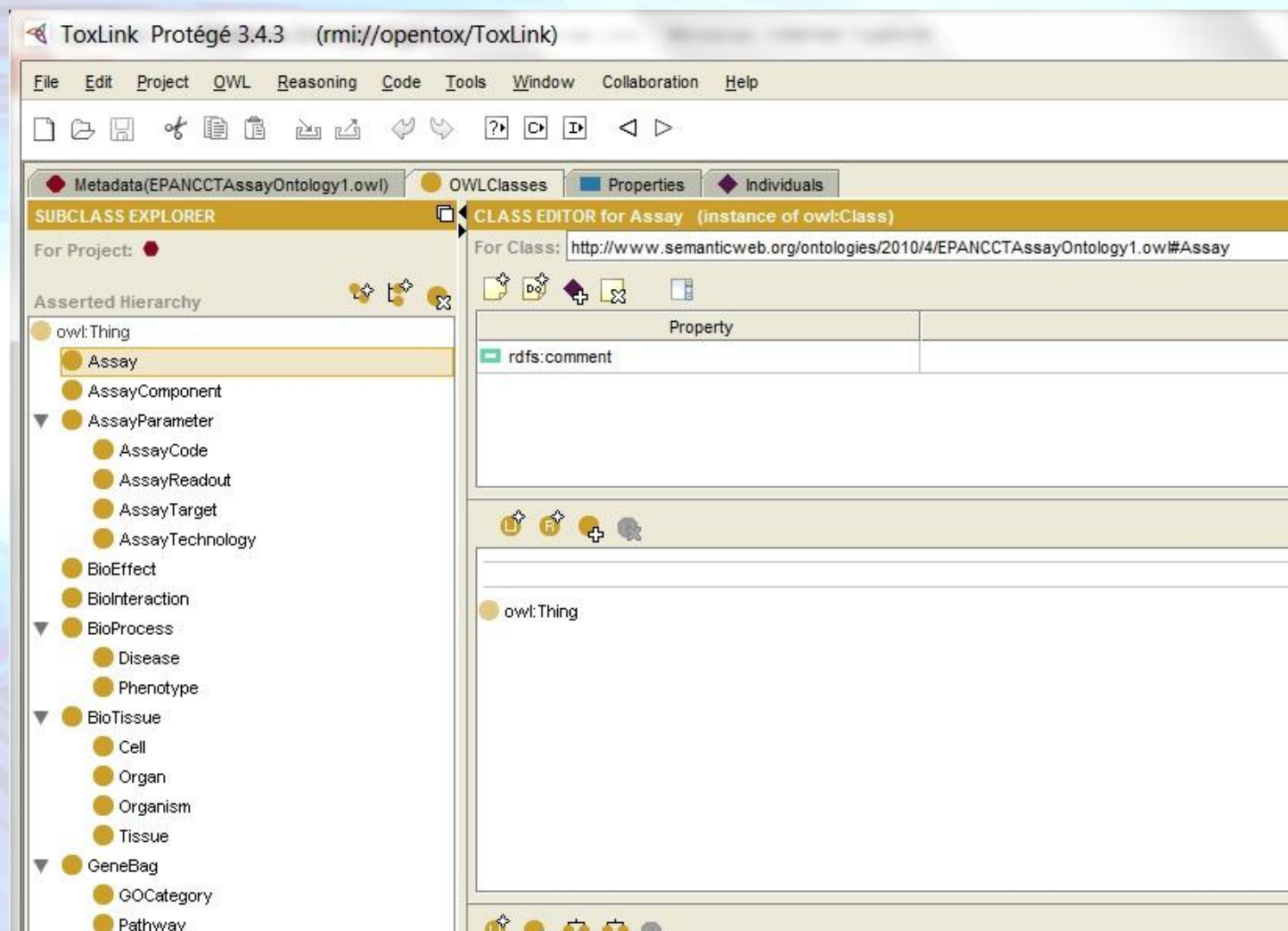


A Toxicology Ontology Roadmap

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), ----

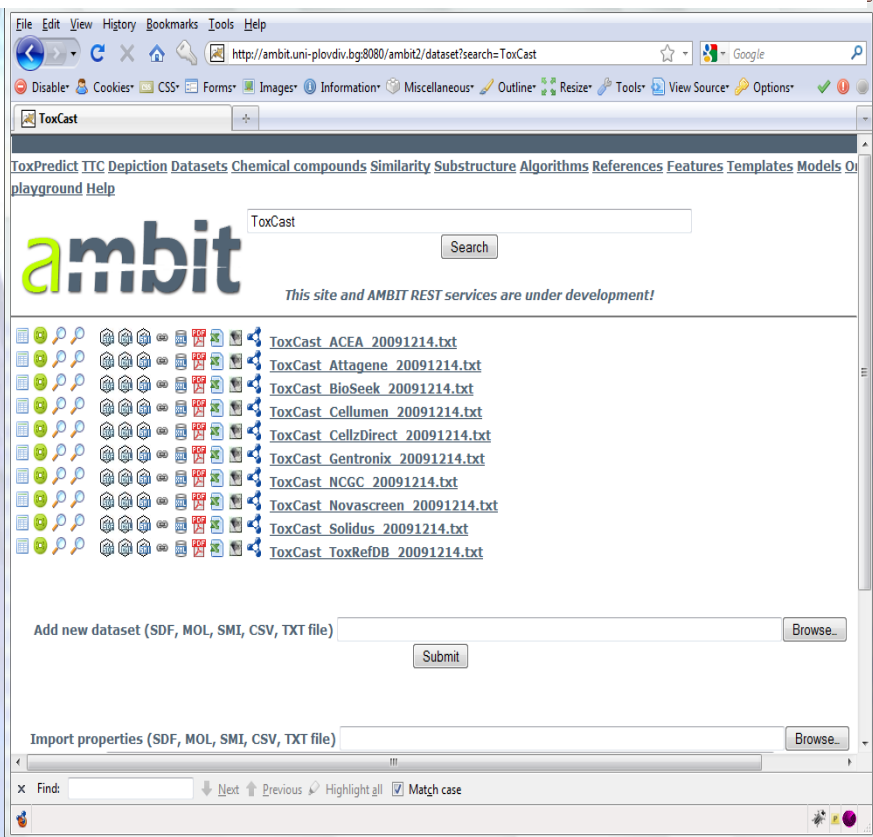
Based on Proceedings from the Toxicology Ontology Roadmap Workshop
EMBL-EBI Industry Programme Workshop
16 -17th November 2010, Hinxton, UK

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/>

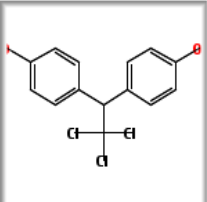
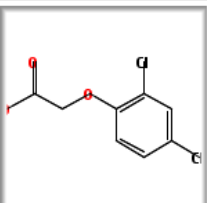
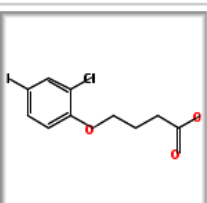

```

    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](#)

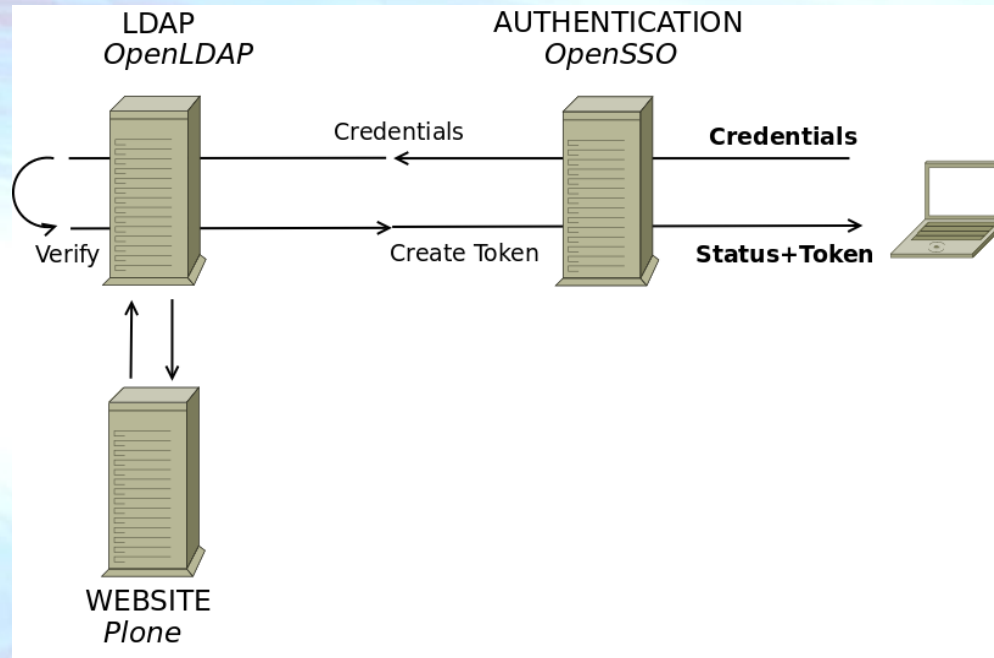
Google

tox.org/toxcas/RORA
tox.org/toxcas/RNR1H2
tox.org/toxcas/RNUT2
tox.org/toxcas/RARA
tox.org/toxcas/RETS1
tox.org/toxcas/RNFKB1
tox.org/toxcas/RNR1D

Controlling Access to Confidential Information

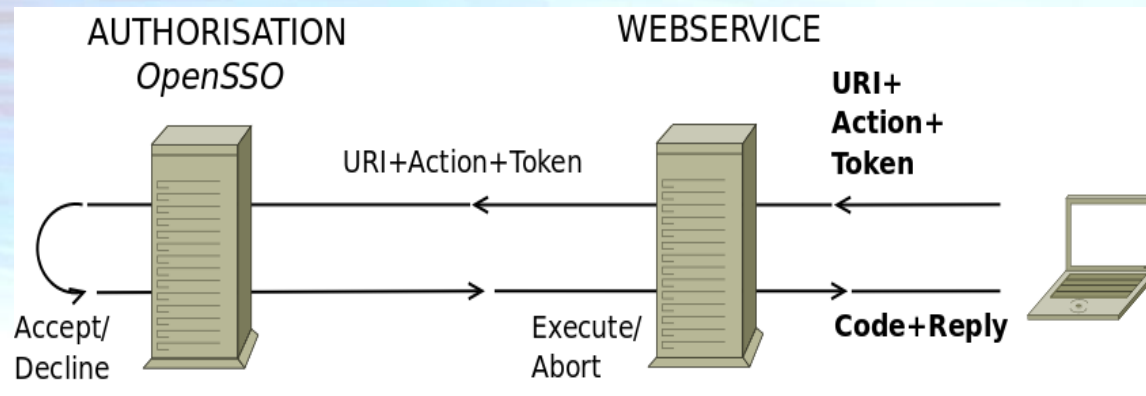
- OpenTox makes resources available through **URIs**
- OpenTox provides facilities to protect confidential information located at **URIs**. Two tasks are involved here:
 - **Authentication**: Confirming the identity of the user requesting access
 - **Authorisation**: Granting the confirmed identity access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



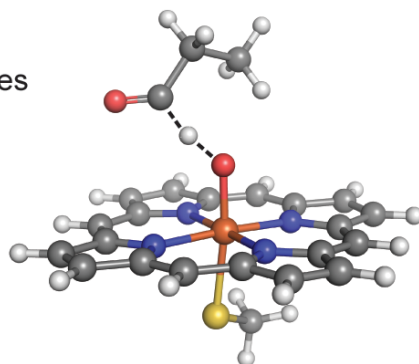
- Tokens encode user identity
- Tokens are valid for a certain time period only (customizable)
- The triplet URI+Action+Token makes up the call to be authorised
- All messages are encrypted (SSL)
- Resource Owners create and modify policies defining access rules

SMARTCyp Service for Predicting Metabolites

Atom Reactivity Library

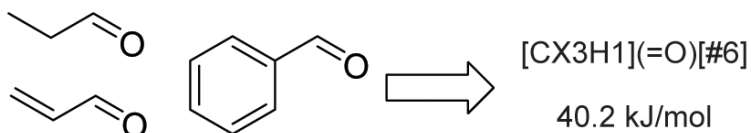
A. Calculate Quantum Chemical Reference Energies

Calculate transition state energies using density functional theory



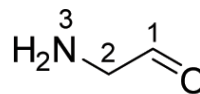
B. Define SMARTS Rules

Group calculations by fragments and calculate average energies



SMARTCyp

1. Assign Energies By SMARTS matching



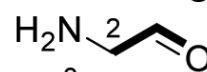
Atom	SMARTS	Energy
1	<chem>[CX3H1](=O)[#6]</chem>	40.2
2	<chem>[CX4][N]</chem>	39.8
3	<chem>[N^3][H1,H2]</chem>	54.1

2. Compute Accessibility Descriptor

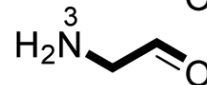
$$A_i = \text{Maxbonds}_i / \text{Maxbonds}_{\text{all}}$$



$$A_1 = 2 / 3 = 0.67$$



$$A_2 = 2 / 3 = 0.67$$



$$A_3 = 3 / 3 = 1.00$$

3. Compute Score and Rank Atoms

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

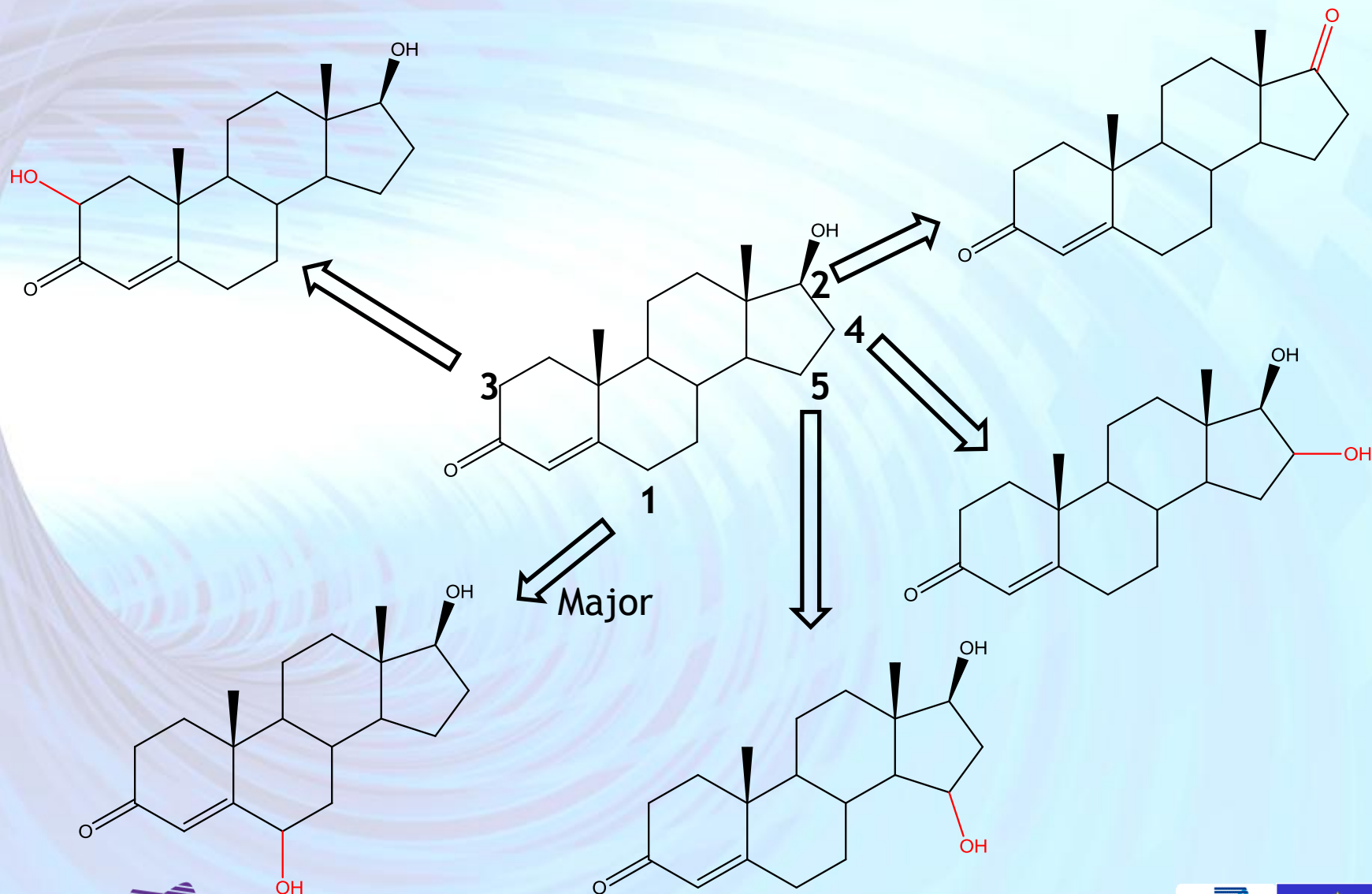
Atom 3 - Rank 3



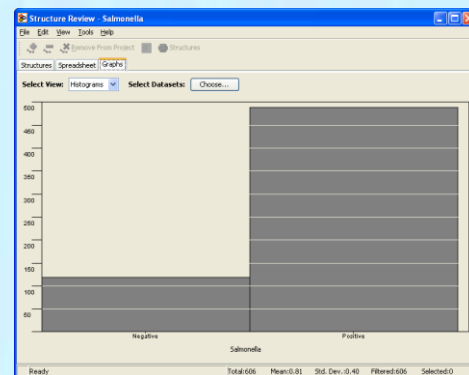
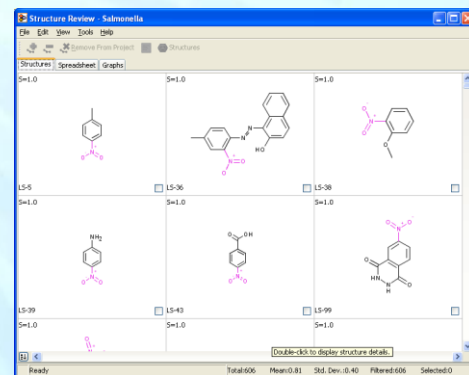
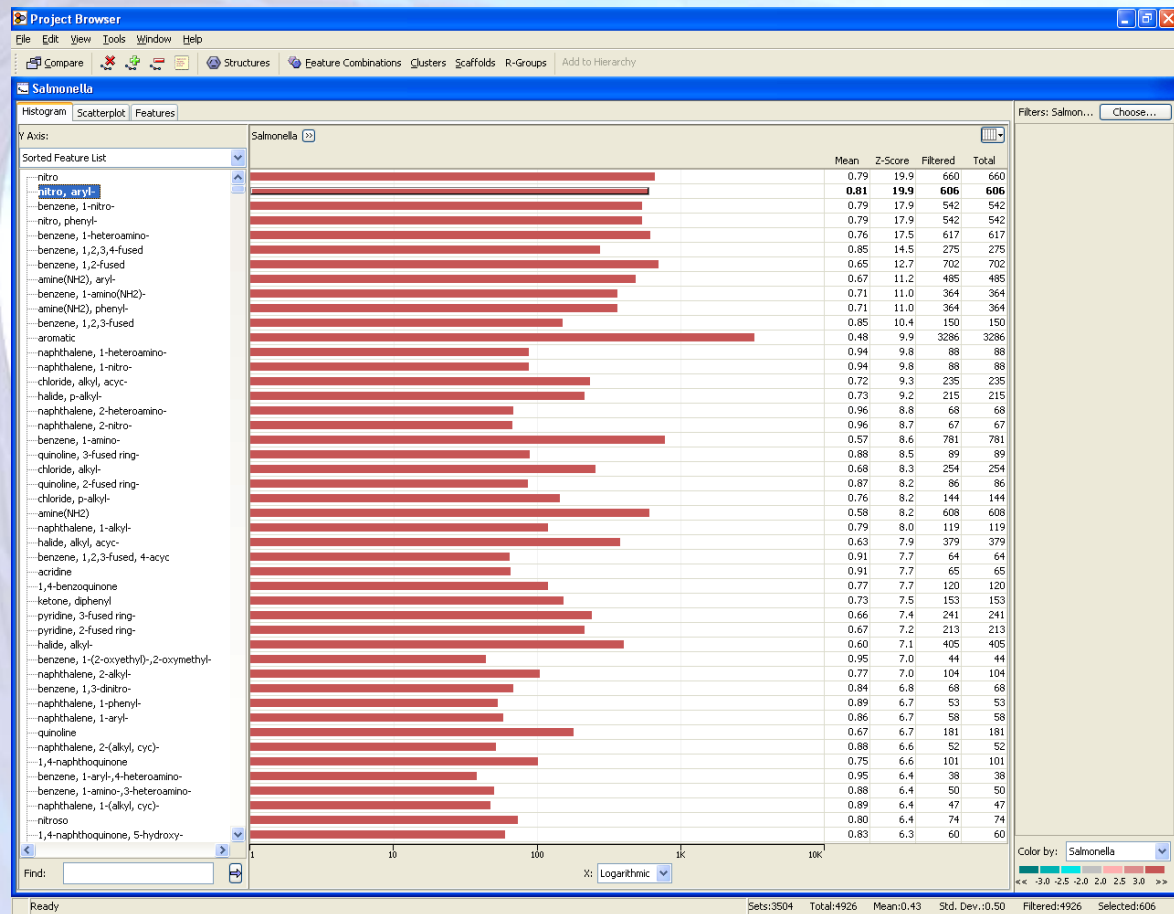
SMARTCyp - developed by Patrik Rydberg, University of Copenhagen

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

SmartCYP Prediction of Testosterone Metabolites



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

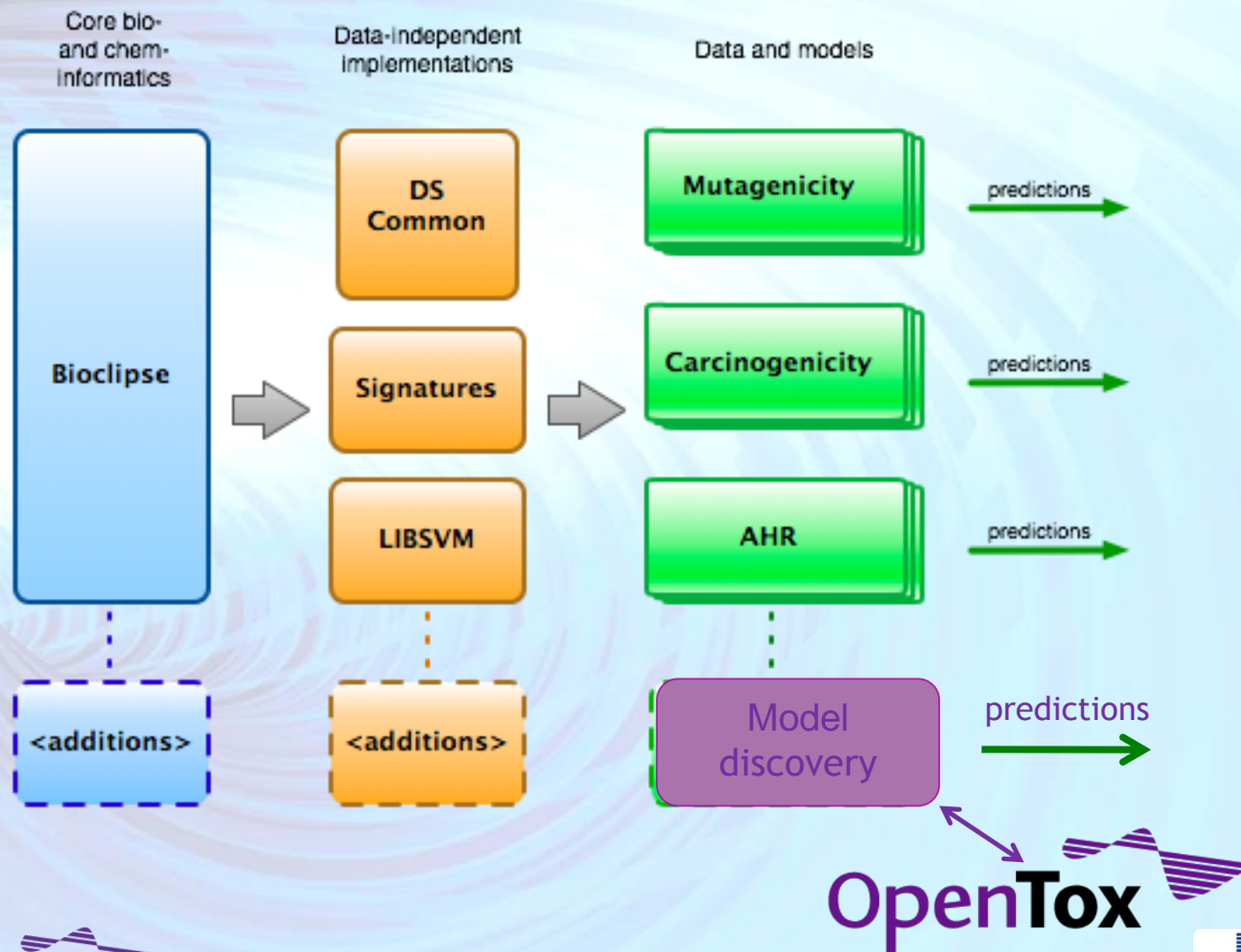
Bioclipse

The screenshot displays the Bioclipse application window. On the left is the 'Bioclipse Navigator' pane showing a project tree with folders like 'CDMS2009', 'ChemGate', 'Gists', 'Media', 'MMSHITDB', 'OWL', 'ADF', 'SampleData', 'Sample Data', 'Solubility', and 'STARLite'. The 'STARLite' folder is expanded, showing files like 'starlite31.sdf', 'Test', 'Test2', 'bug824.js', 'foo.js', 'Fragments2_3d.sdf', 'glt-22185.js', 'unnamed.cml', 'unnamed1.cml', and 'unnamed2.cml'. The main workspace shows a table with two rows of chemical structures. The first row has a molecular structure, the number '14577' in the first column, and '17159' in the third column. The second row has a similar molecular structure, the number '14578' in the first column, and '17160' in the third column. On the right is the 'Properties' pane, which is currently showing the 'General' tab. It lists properties such as 'Has 2D Coords' (yes), 'Has 3D Coords' (no), 'Molecular Format' (N/A), 'Molecular Formula' (C18H26N6O6P), 'Molecular Mass' (358.1781), and 'Molecular Properties' (MOLREGNO: 17159).

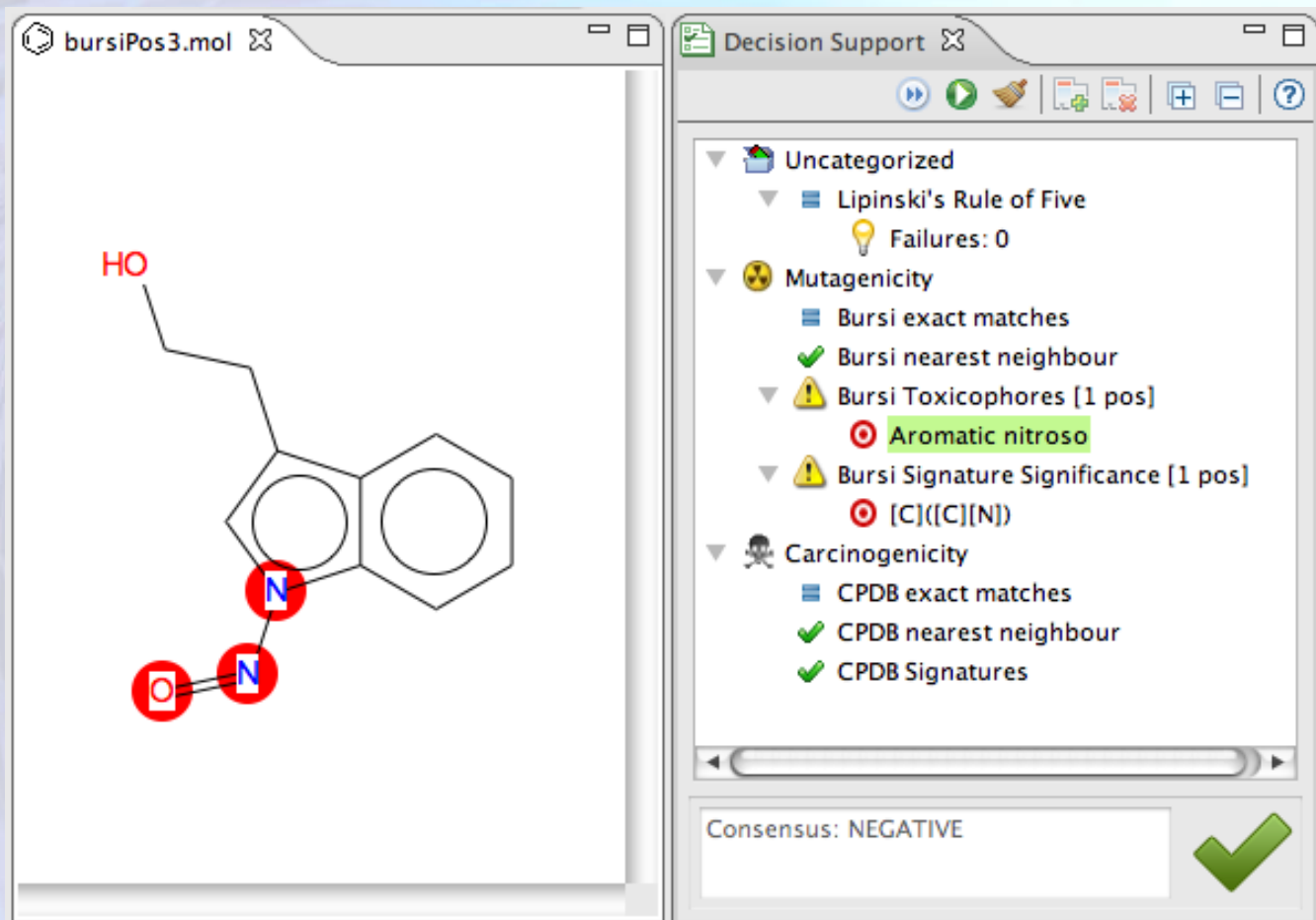
	2D-structure	MOLREGNO
14577		17159
14578		17160

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

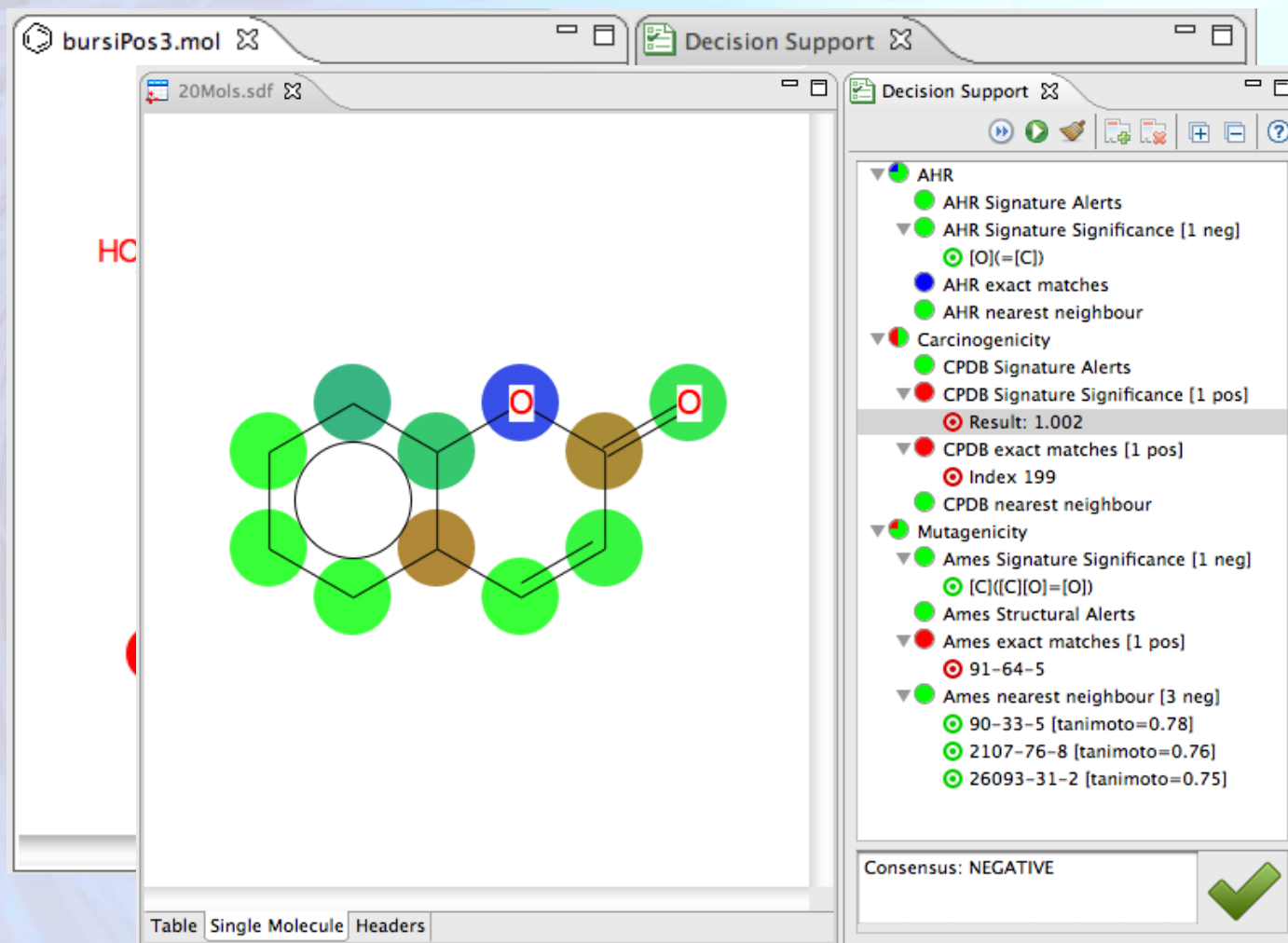
Bioclipse - OpenTox Interoperation



Bioclipse Visualisation Workbench



Bioclipse Visualisation Workbench



Bioclipse Visualisation Workbench

HC

20Mols.sdf

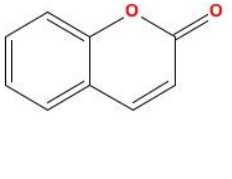
Report Editor

Decision Support

Decision Sup... Default

Report date: 2010/08/26

Query structure:



Properties

Mol.	146.143
H donors:	1
H	1
alogP:	1.031

Number of Consensus: 4
Endpoints: 2 negative, 1 inconclusive

Compound: coumarin
SMILES: O=C1OC2=CC=CC=C2C=C1
Formula: C₉H₆O₂
InChI: InChI=1S/C9H6O2/c10-9-6-5-7-3-1-2-4-8(7)/11-9/h1-6H

Endpoint: AHR **NEGATIVE**

Model: AHR exact matches
Consensus: INCONCLUSIVE

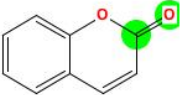
Model: AHR nearest neighbour
Consensus: NEGATIVE

Model: AHR Signature Alerts
Consensus: NEGATIVE

Model: AHR Signature Significance
Consensus: NEGATIVE

Compound: [O]([C])
Classification: **NEGATIVE**

Details:



1 of 3

Decision Support

- AHR
 - AHR Signature Alerts
 - AHR Signature Significance [1 neg]
 - [O]([C])
 - AHR exact matches
 - AHR nearest neighbour
- Carcinogenicity
 - CPDB Signature Alerts
 - CPDB Signature Significance [1 pos]
 - Result: 1.002
 - CPDB exact matches [1 pos]
 - Index 199
 - CPDB nearest neighbour
- Mutagenicity
 - Ames Signature Significance [1 neg]
 - [C]([C])(O)=[O]
 - Ames Structural Alerts
 - Ames exact matches [1 pos]
 - 91-64-5
 - Ames nearest neighbour [3 neg]
 - 90-33-5 [tanimoto=0.78]
 - 2107-76-8 [tanimoto=0.76]
 - 26093-31-2 [tanimoto=0.75]

Consensus: NEGATIVE

OpenTox - ToxCast

ACToR: Aggregated Computational Toxicology Resource

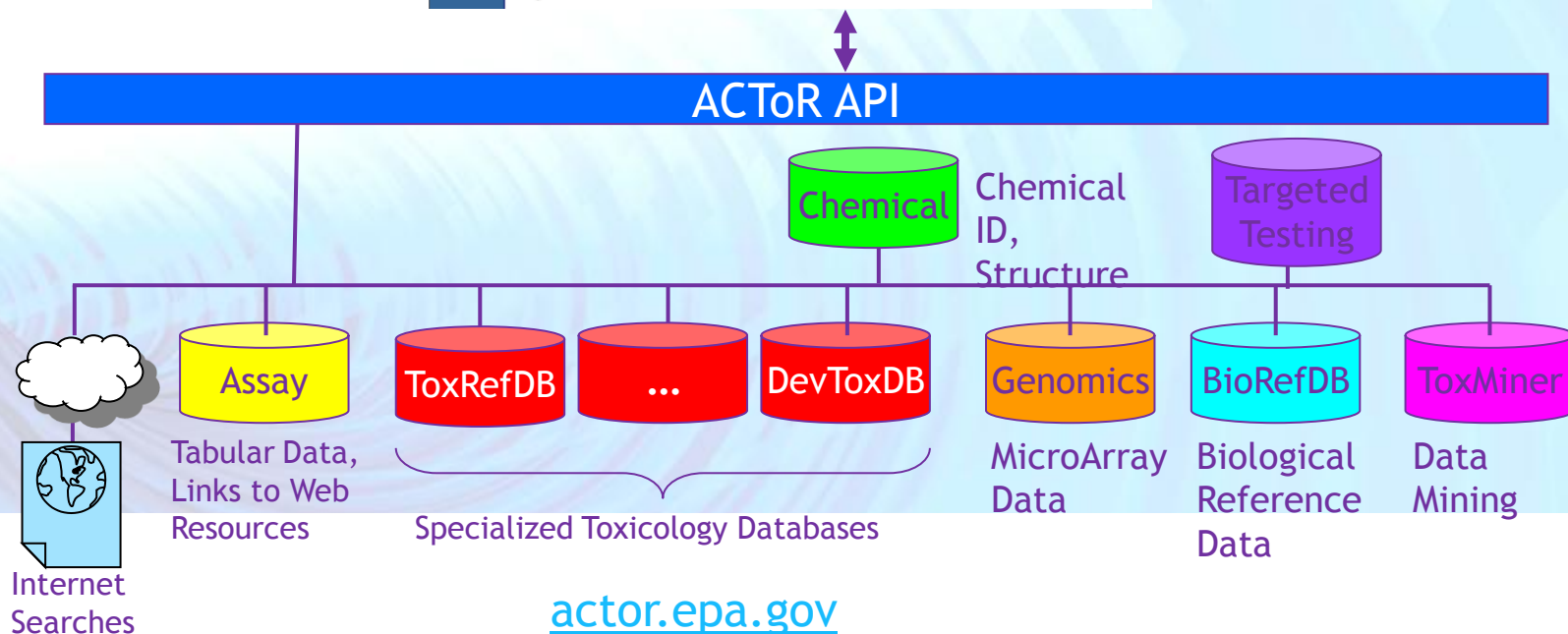
Data Collection: EPA CCL3

Name: EPA CCL3 List.pdf
 Description: EPA has drinking water regulations for more than 90 contaminants. The Safe Drinking Water Act (SDWA) includes a process that we must follow to identify and list unregulated contaminants which may require a national drinking water regulation in the future. EPA must periodically publish this list of contaminants (called the Contaminant Candidate List or CCL). In February 2008 we announced the draft CCL 3.
 ID: 139
 Institutional Source: EPA
 Source Type: Chemicals
 Number of Substances: 93
 Number of Generic Chemicals: 92

Chemical Table
 Page 1 of 2

Structure	Name	CASRN	Generic Chemical Details	Hazard	Chronic Toxicity	Developmental Toxicity	Genotoxicity	Reproductive Toxicity	Chemical Toxicity	Food Safety
<chem>ClC(Cl)(Cl)Cl</chem>	1,1,1,2-Tetrachloroethane	630-20-6	Details	H ₃₅₀	C ₃₅₀	G ₃₅₀	D ₃₅₀	R ₃₅₀	Cy ₃₅₀	
<chem>ClCCl</chem>	1,1-Dichloroethane	75-34-3	Details	H ₃₅₀	C ₃₅₀	G ₃₅₀	D ₃₅₀	R ₃₅₀	Cy ₃₅₀	

ACToR Web Browser



OpenTox - ToxCast

ToxPredict

OpenTox demo application

Step 1: Search
Select structure(s)

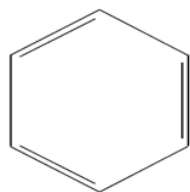
Step 2: Verify structure
Verify structure

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

Download as

MolecularWeight **MolecularWeight**
MW 78.1112



Internet Searches

Tabular Data,
Links to Web
Resources

Specialized Toxicology Databases

actor.epa.gov

U.S. ENVIRONMENTAL PROTECTION AGENCY

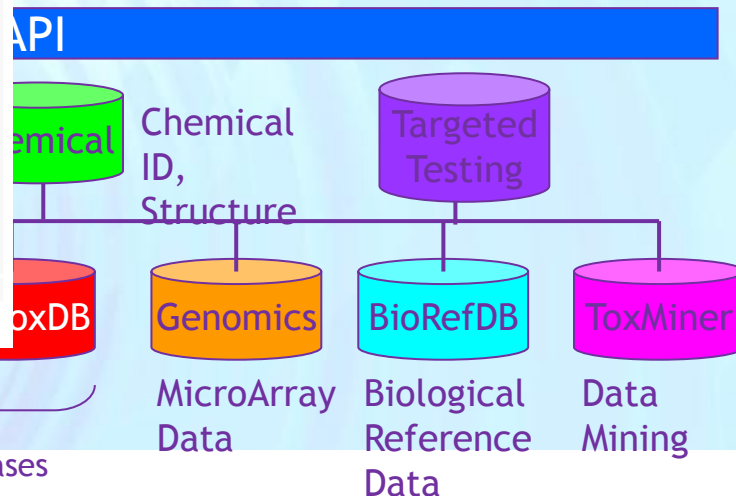
Welcome, **guest**
[Help](#)
[Admin](#)

Under Act (SDWA) includes a process that we must follow to identify and list unregulated or periodically publish this list of contaminants (called the Contaminant Candidate List or CCL).

Health Safety
 Chemical Safety
 Regulatory Toxicology
 Environmental Health
 Research
 Data
 Hazard

C4 **G** **D** **R** **O**
C2 **G** **D** **R** **O**

ACToR Web
Browser



ToxBank Infrastructure Project

(started Jan 2011)

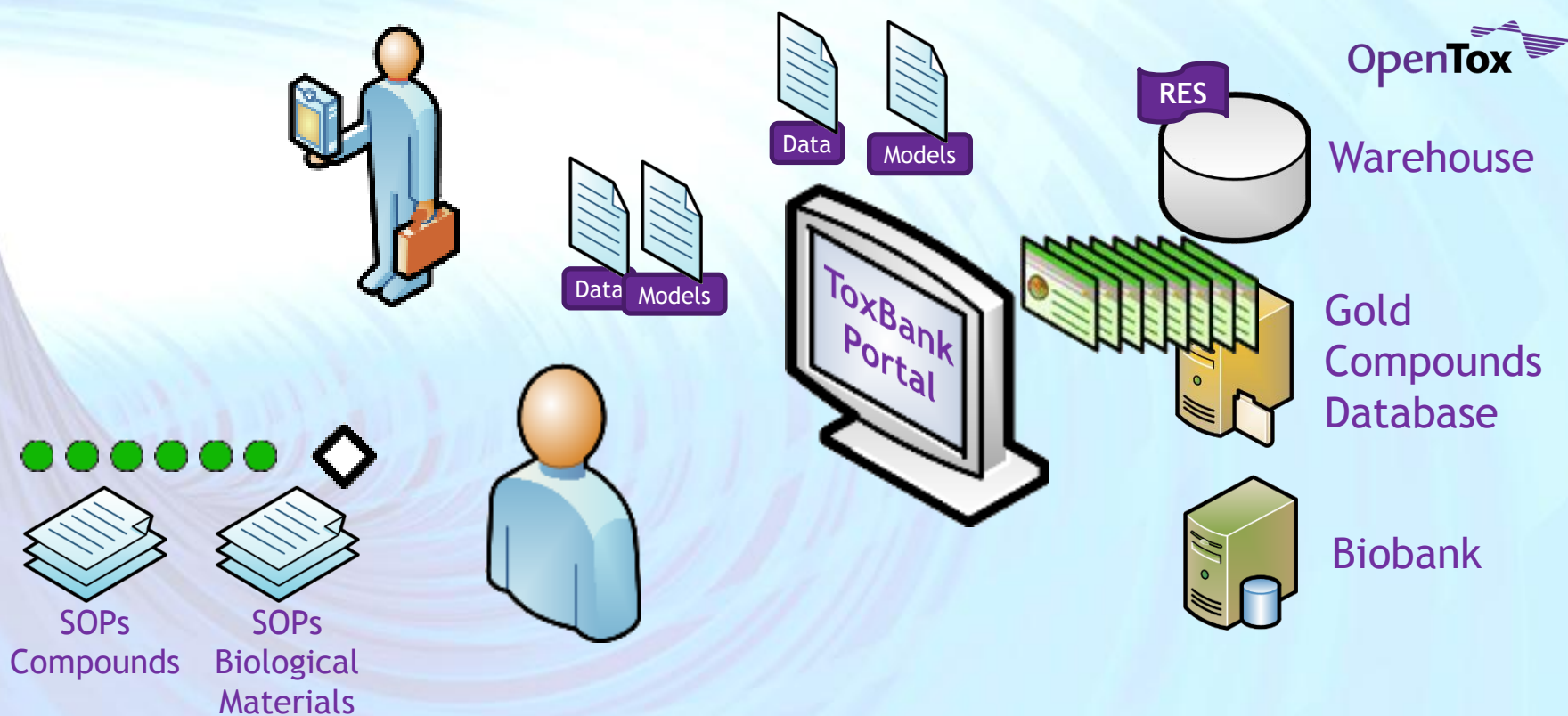
Establishment of a ...



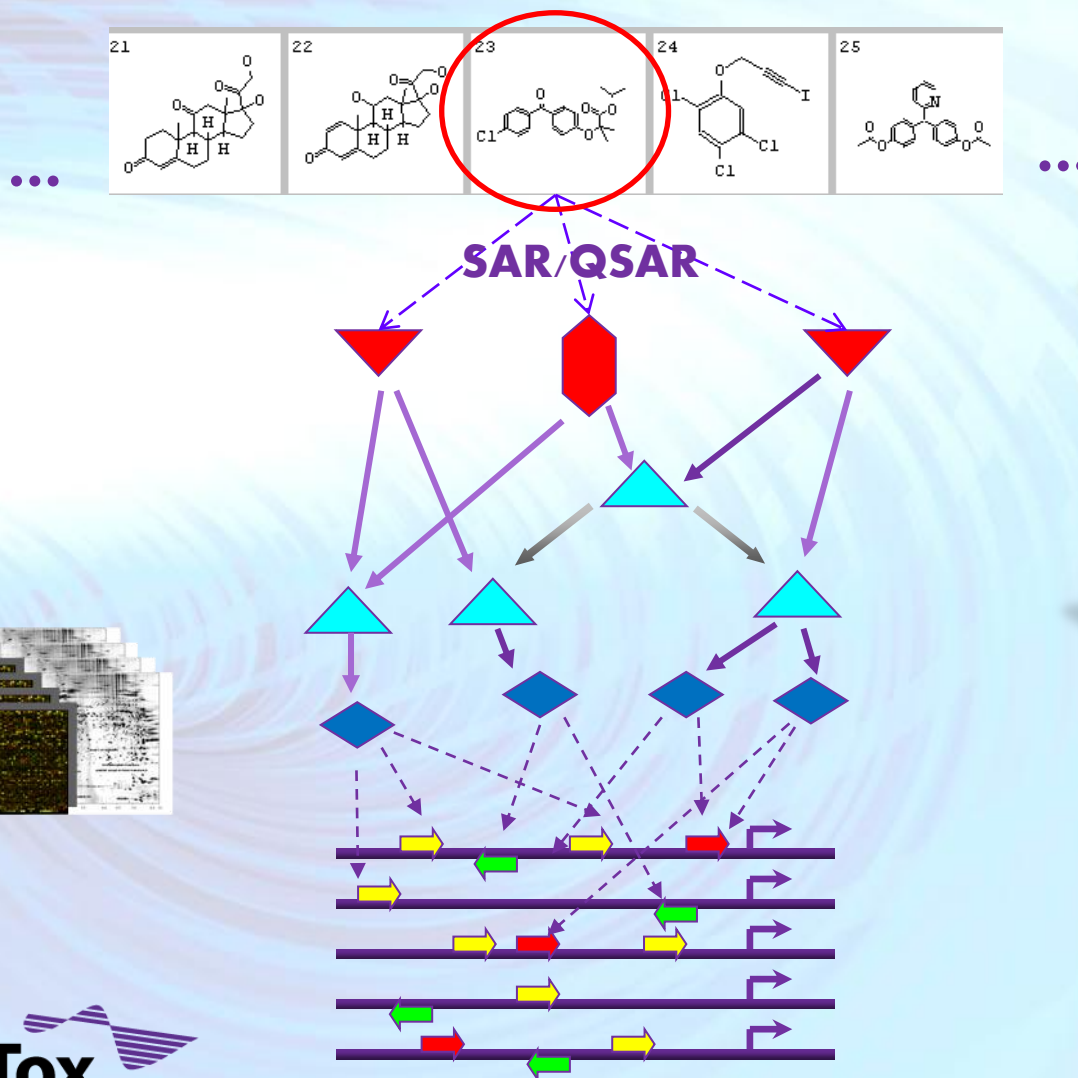
- ... cell and tissue banking information resource
- ... repository for the selected test compounds
- ... database of reference test compounds
- ... dedicated web-based data warehouse

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



Integrated Workflow of Bioinformatics, Systems Biological and Cheminformatics Tools



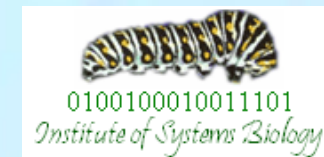
Cheminformatics
Biological activities of
the compounds

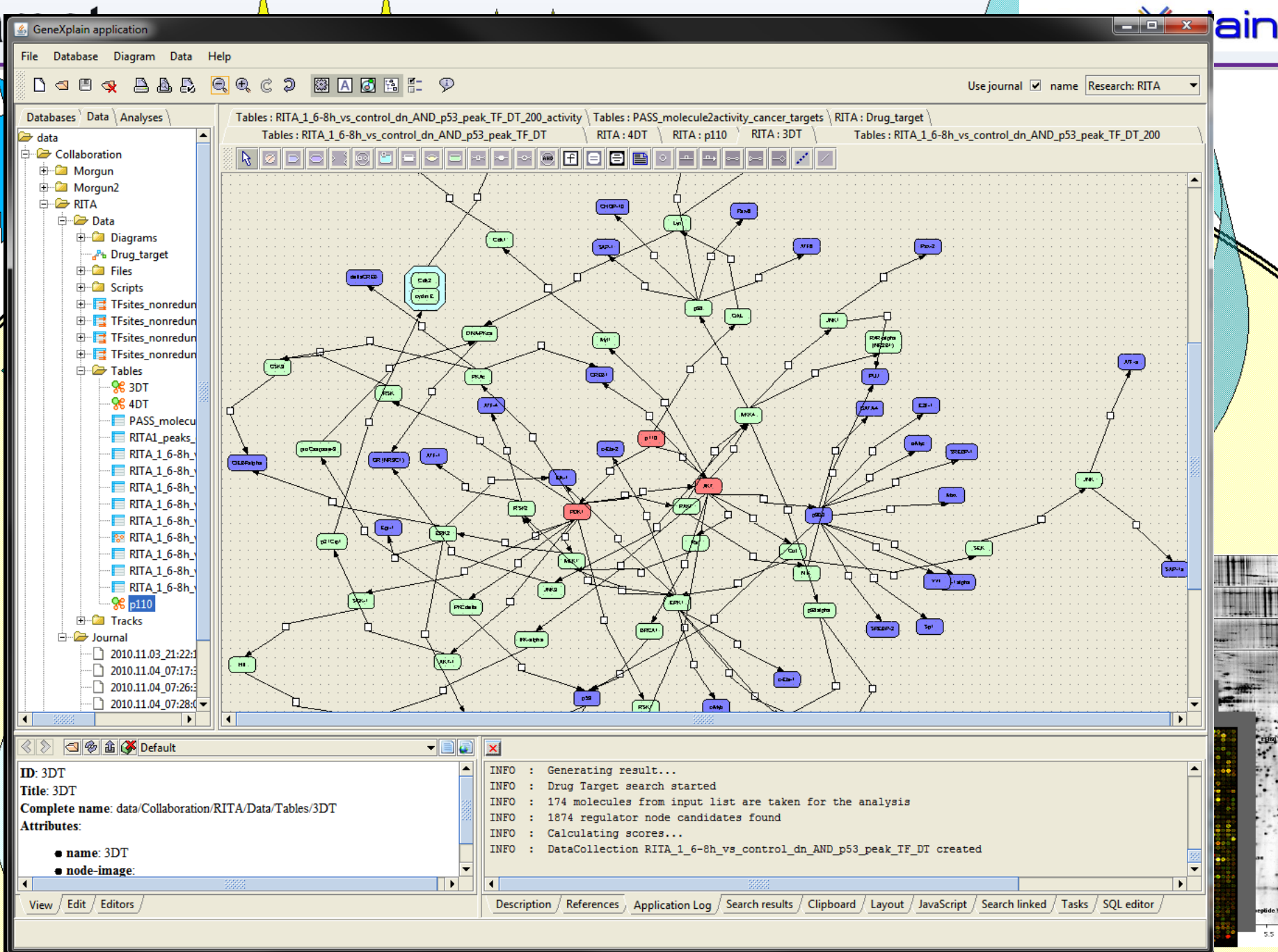
Systems Biology
Network analysis and
simulation, potential
drug targets

Bioinformatics
Regulatory modules,
transcription factors

Communicating Biological models via BioUML

Unified Modeling Language (UML) is a
standardized general-purpose modeling
language



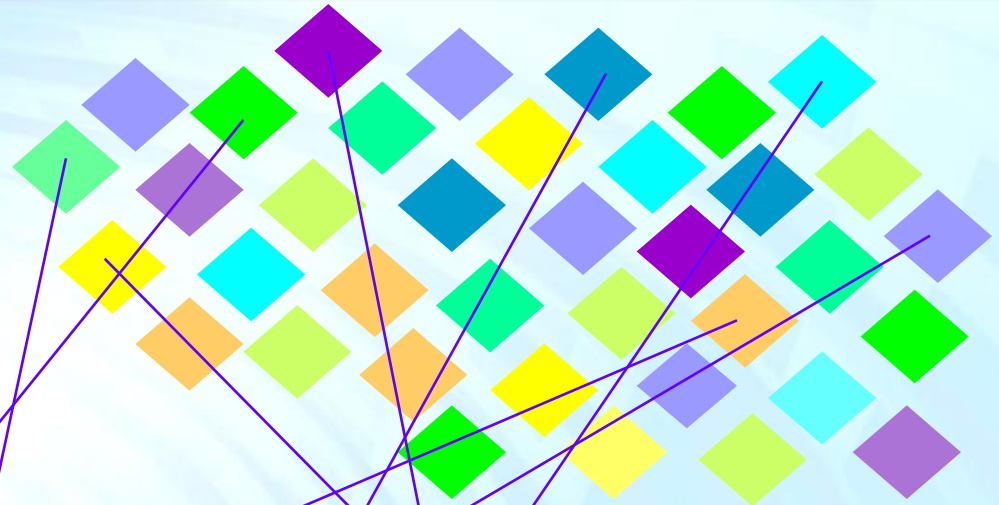


Creation of VO from Collaboration Pool

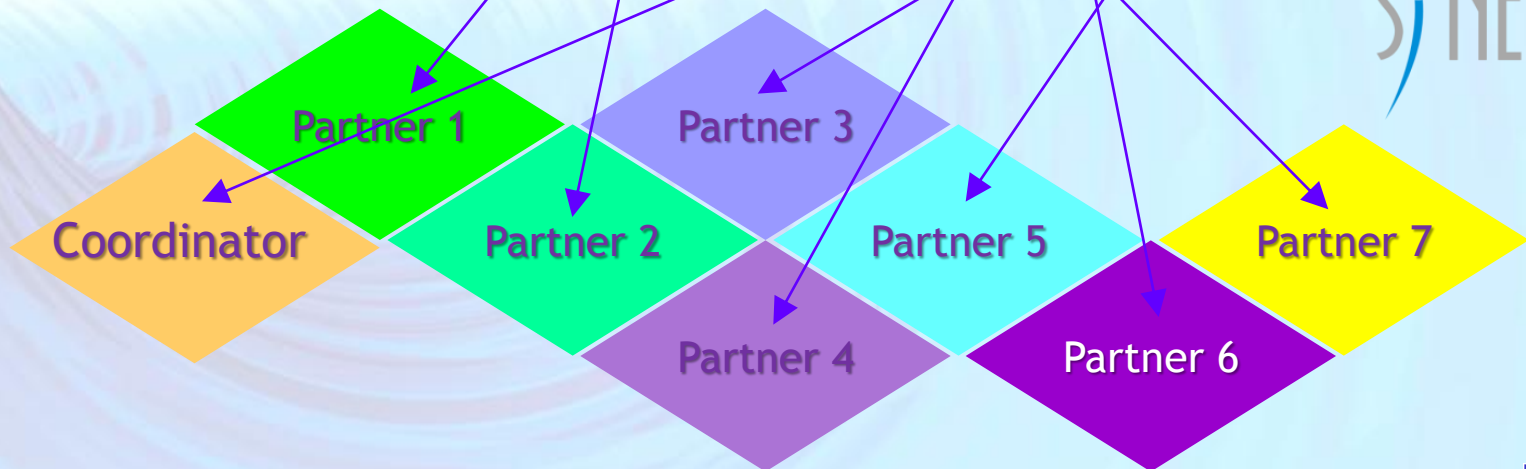
Network

Opportunity

Call for Tender
Need for joint effort
Major project

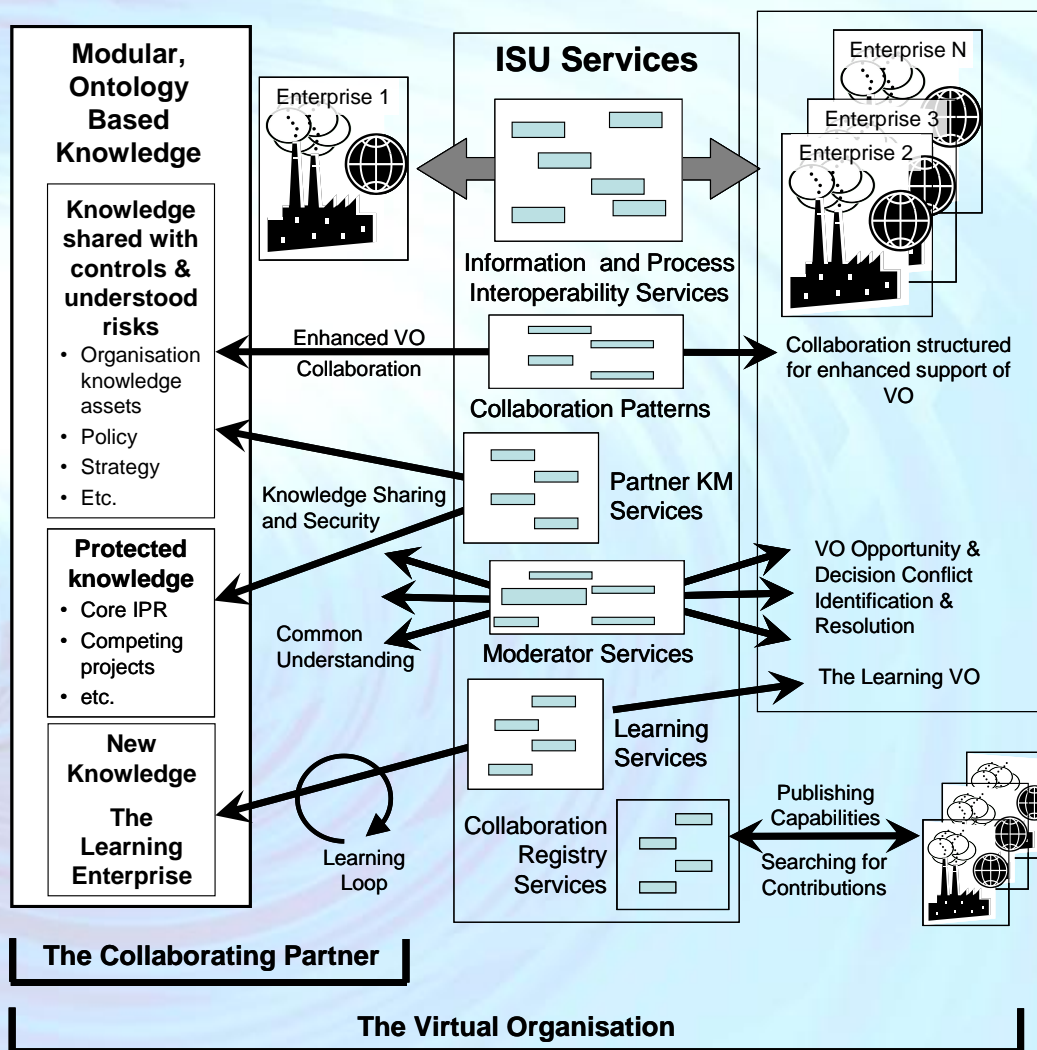


Virtual Organisation

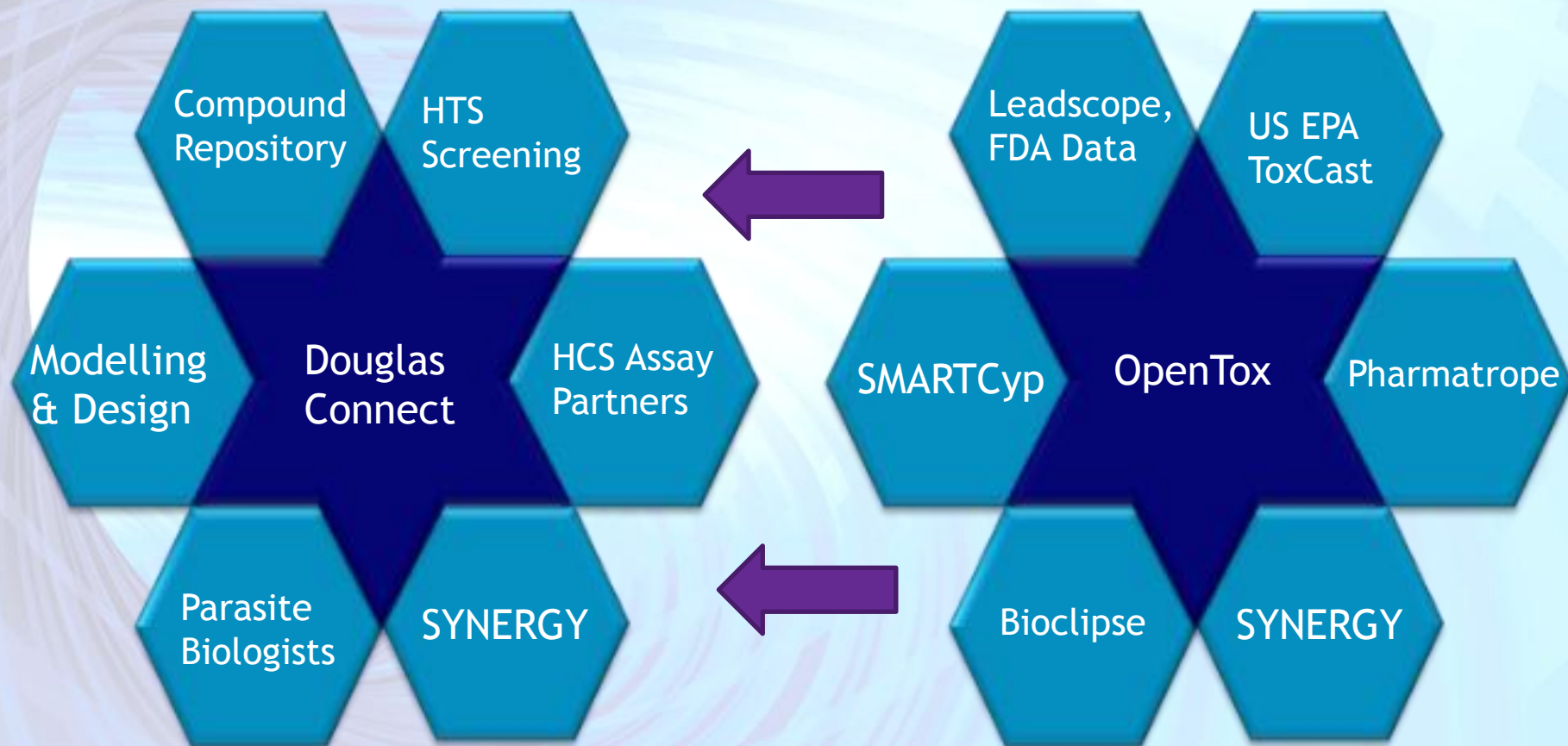


SYNERGY

SYNERGY Collaboration Services



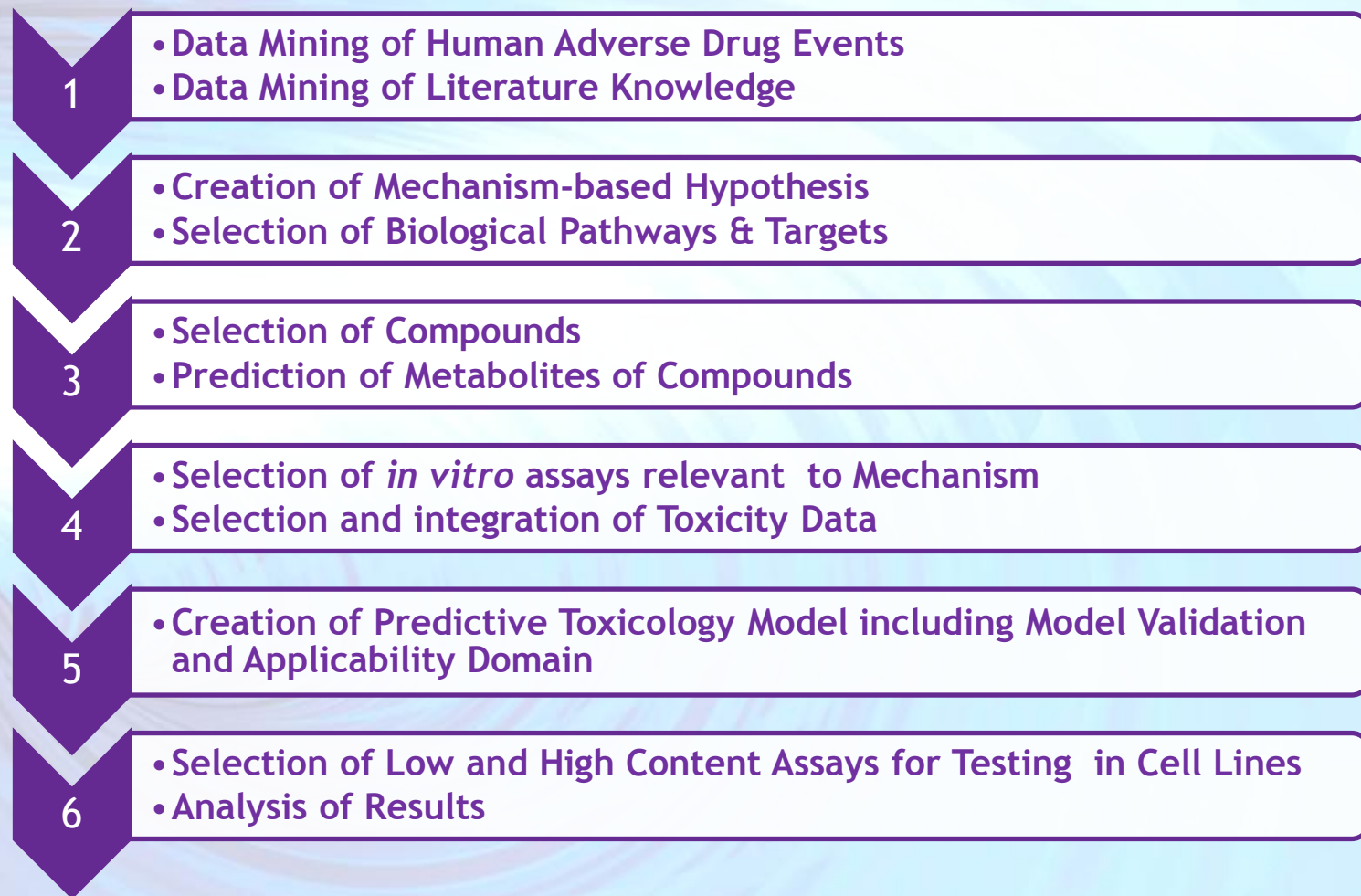
Virtual Organisation Pilots



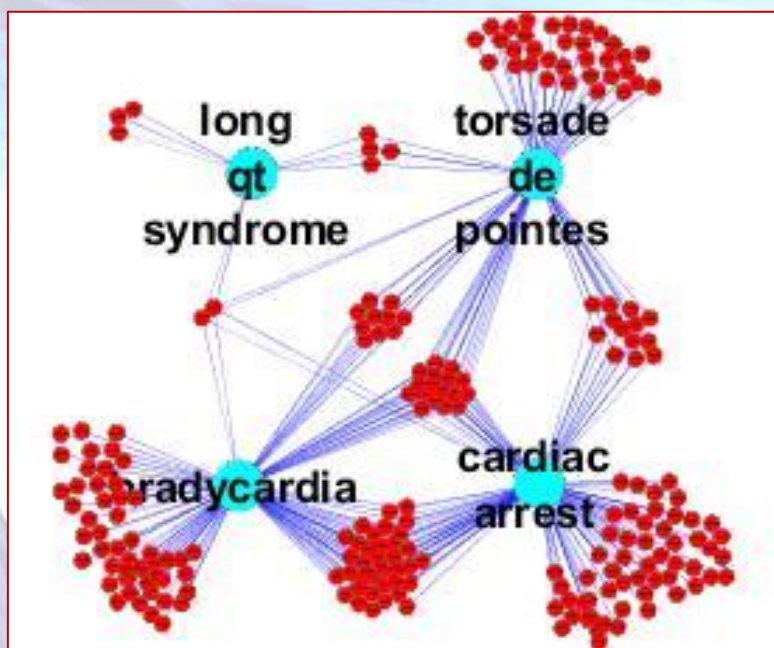
Neglected Disease Drug Design VO

Predictive Toxicology VO

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



Analysis of Adverse Events Based on Pharmacological Activity



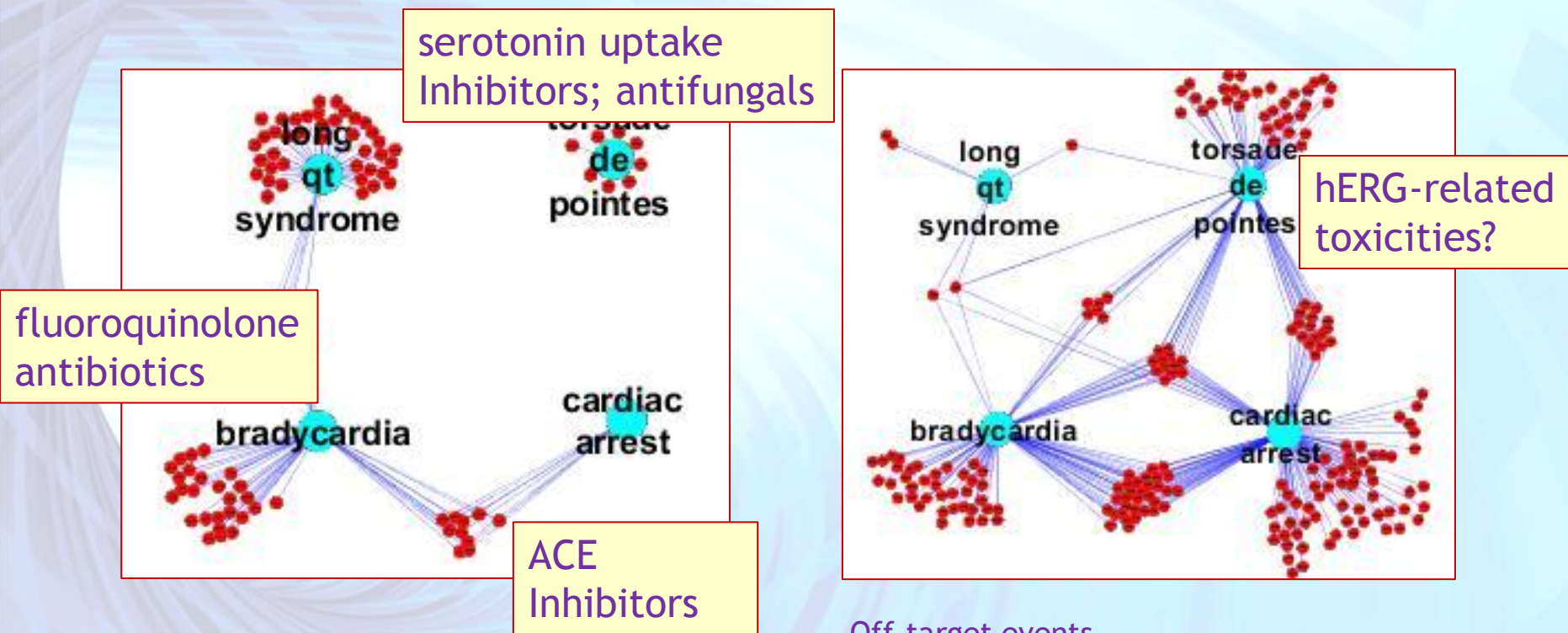
- Cardiac adverse events
- Related to hERG ion channel?

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?



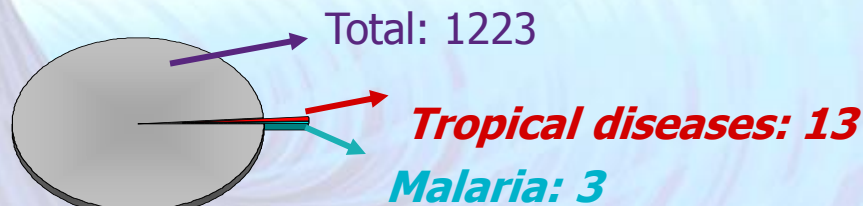
Example: Cardiac Adverse Events



PHARMATROPE

Neglected Diseases - Malaria

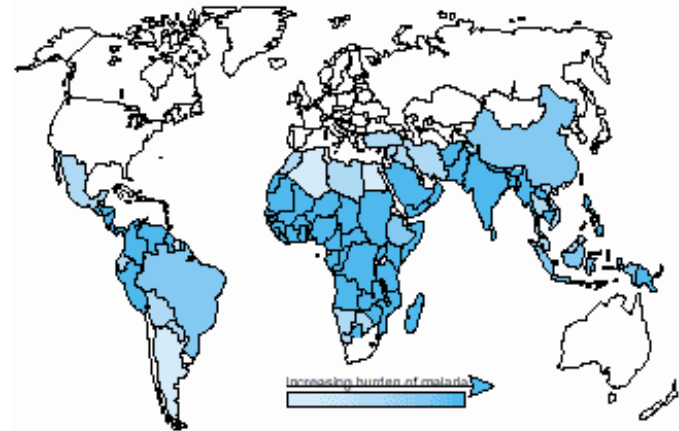
- 2.5 billion people at risk
- 500 M cases yearly
- 1-3 M deaths yearly
- Many child fatalities
- Brain Damage, Impaired Development
- Few drugs, no vaccine yet



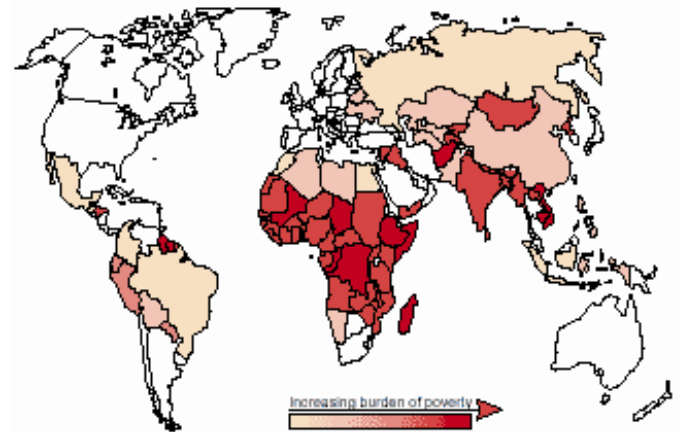
*Drug development outcome,
last quarter of the XXth
century*

Greenwood & Mutabingwa, Nature 415:670-672

Estimate of world malaria burden

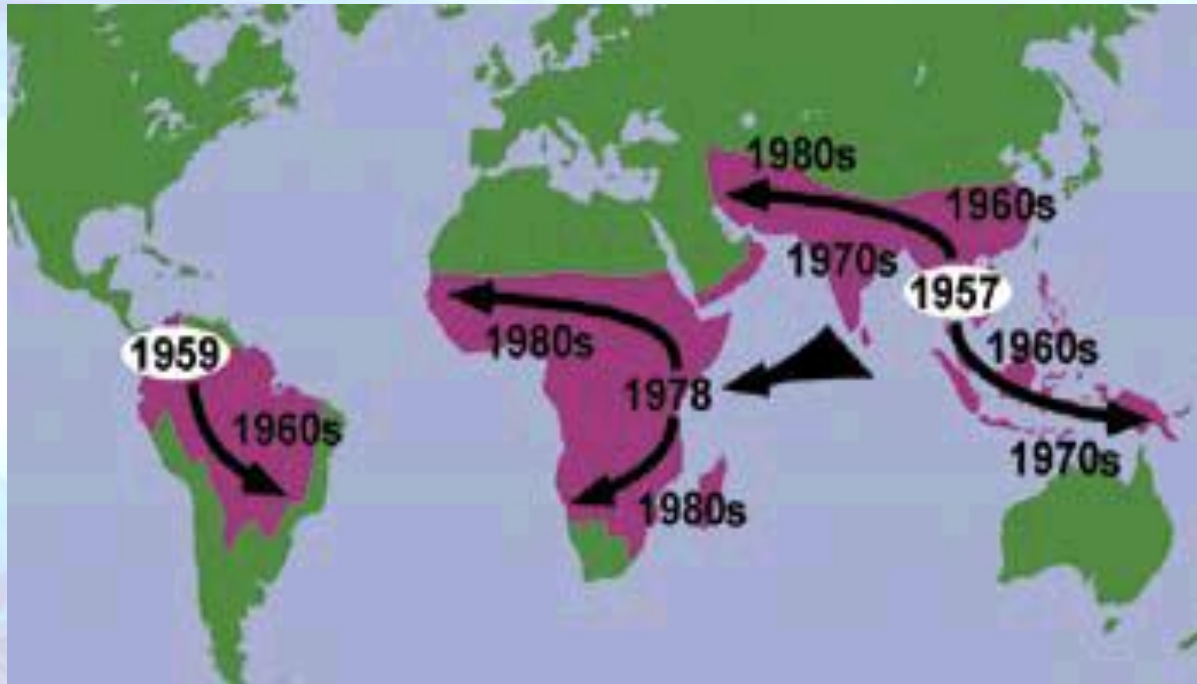


Estimate of world poverty



Source: RBM data/J. Sachs 1999

Anti-malarial Drug Resistance



Emergence and spread of chloroquine resistance

Drug Design VO Roles

The VO involves the following partner roles:

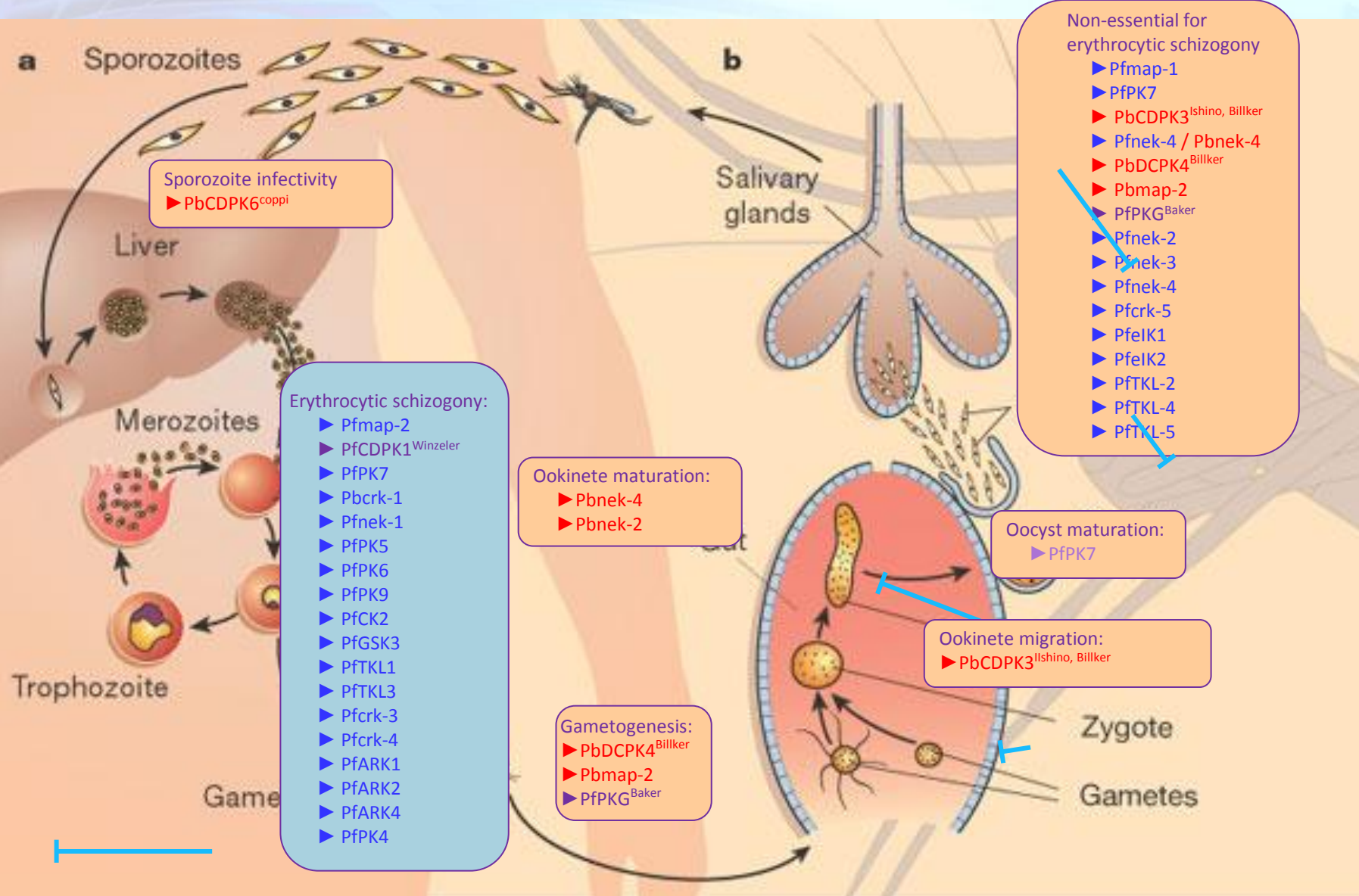
1. Project Management (PM)
2. Crystallographer (CRYS)
3. Pharmacophore (PHORE)
4. Virtual Screening (VS)
5. Docking (DOCK)
6. Predictive ADME (ADME)
7. Predictive Toxicology (TOX)
8. Library Design (LIB)
9. Compound Supplier (CMPD)
10. Protein Supplier (PROT)
11. Activity Assay (AA)
12. Toxicology Assay (TA)
13. Information and Communications Technology (ICT)

Scientists Against Malaria (SAM)

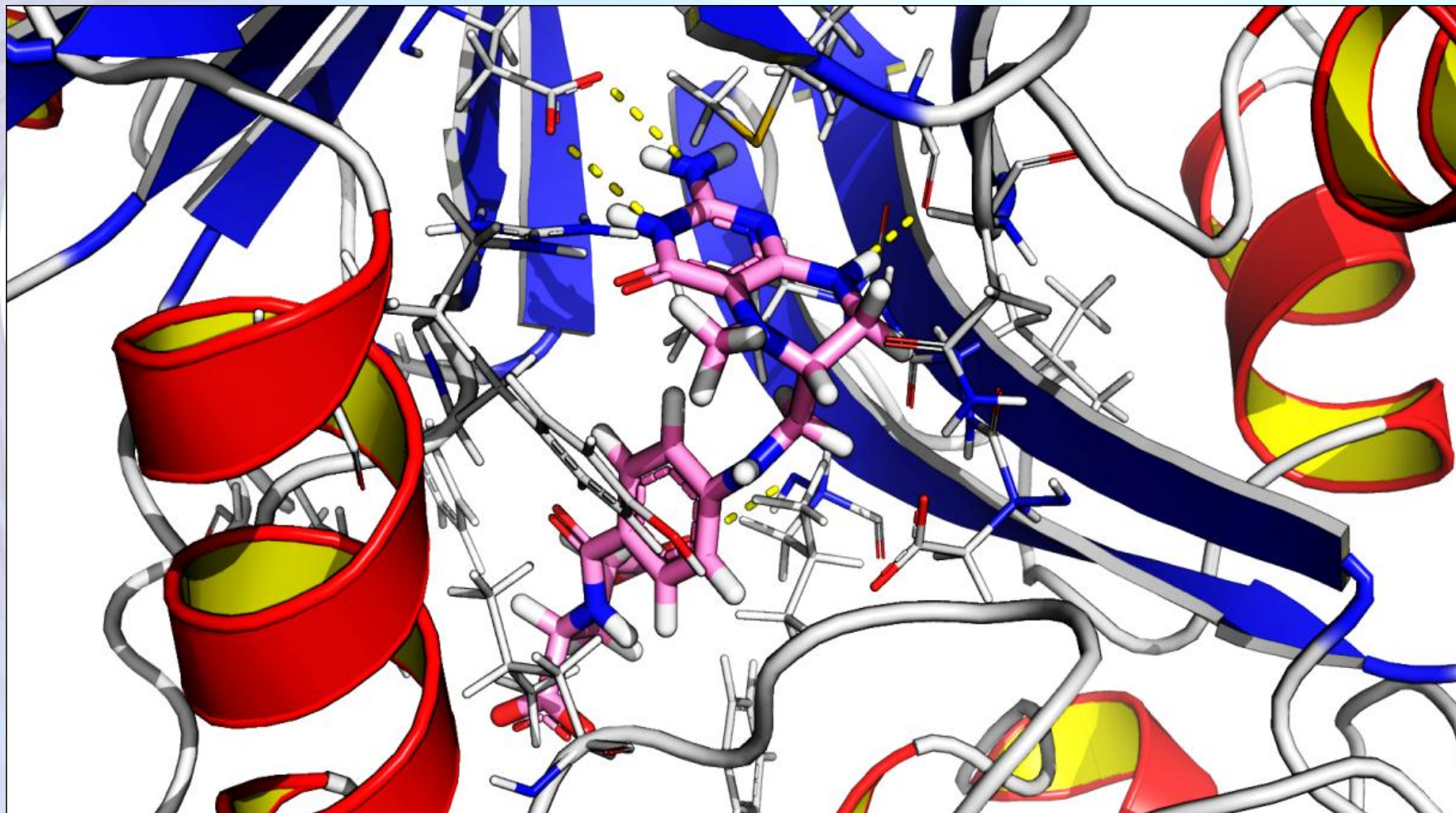


Investigators collaborating on the research include Barry Hardy & Roman Affentranger (Douglas Connect), Alessandro Contini (University of Milan), Hugo Gutierrez de Teran (Public Galician Foundation of Genomic Medicine), Jeffrey Wiseman & Matt Clark (Pharmatropé), Jeff Spitzner (Rescentris), Ruben Papoian, William Seibel & Sandra Nelson (University of Cincinnati Drug Discovery Center), Sharon Bryant (Inte:Ligand), Andrew Wilks & Isabelle Lucet (Monash University) & Christian Doerig, Coordinator of the FP7 MALSIG project on signalling in Malarial parasites, and Matteo Dal Peraro (EPFL, Lausanne).

SAM VO targeting Plasmodium Kinases

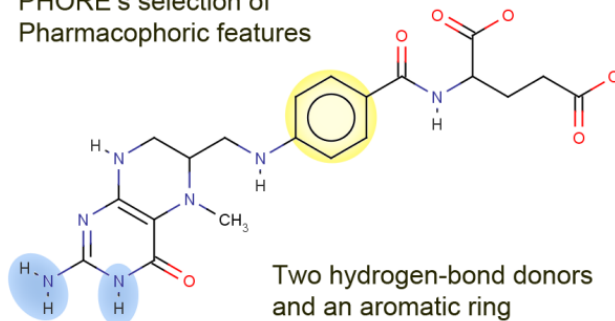


Target-based Drug Design approach

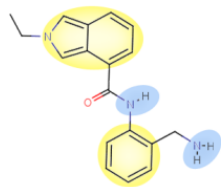


Virtual Screening

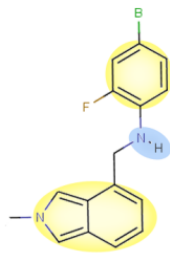
PHORE's selection of
Pharmacophoric features



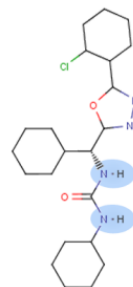
PHORE's library of compounds



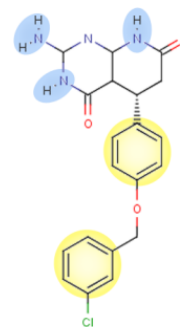
Compound-ID 001
aromatic ring
2 donors



Compound-ID 002
Only one donor



Compound-ID 003
No aromatic ring

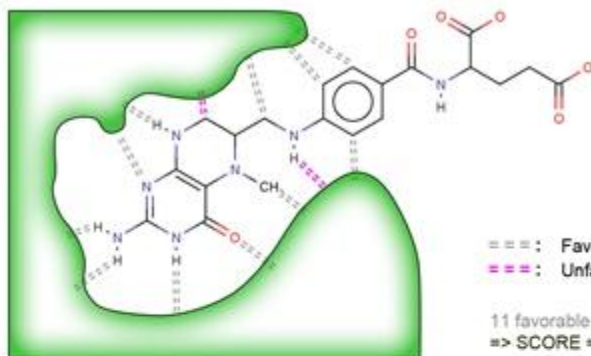


Compound-ID 004
aromatic ring
2 donors

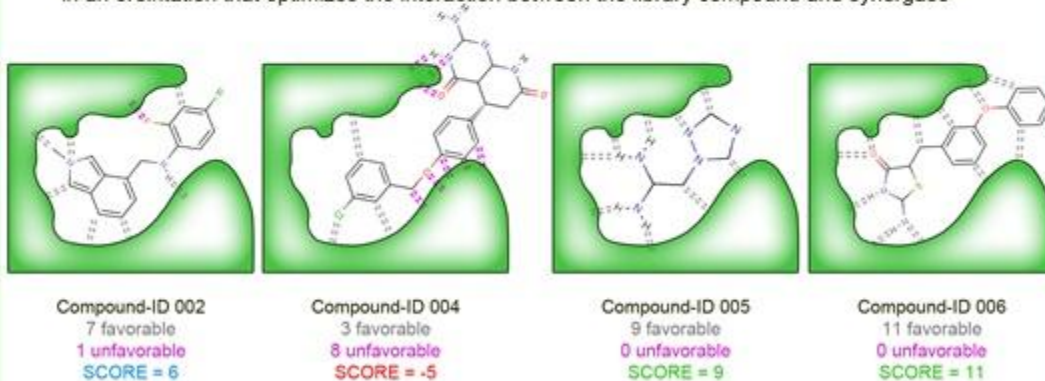
PHORE's activity hits prediction: compound-IDs 001 and 004

Docking

γ -synergine bound to synergase (schematic depiction)

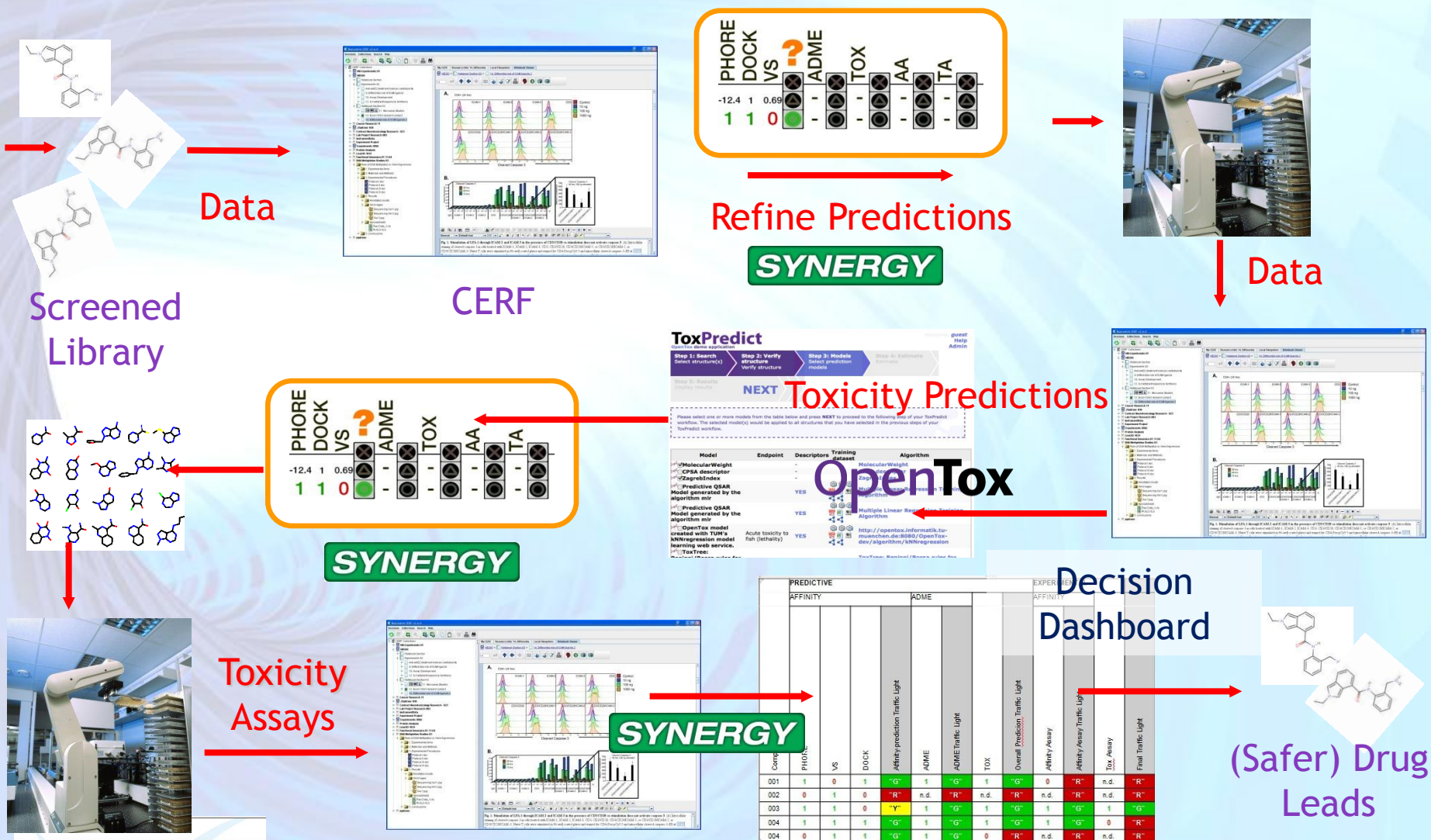


Each member of DOCK's library of compounds is placed in the binding pocket of synergase in an orientation that optimizes the interaction between the library compound and synergase

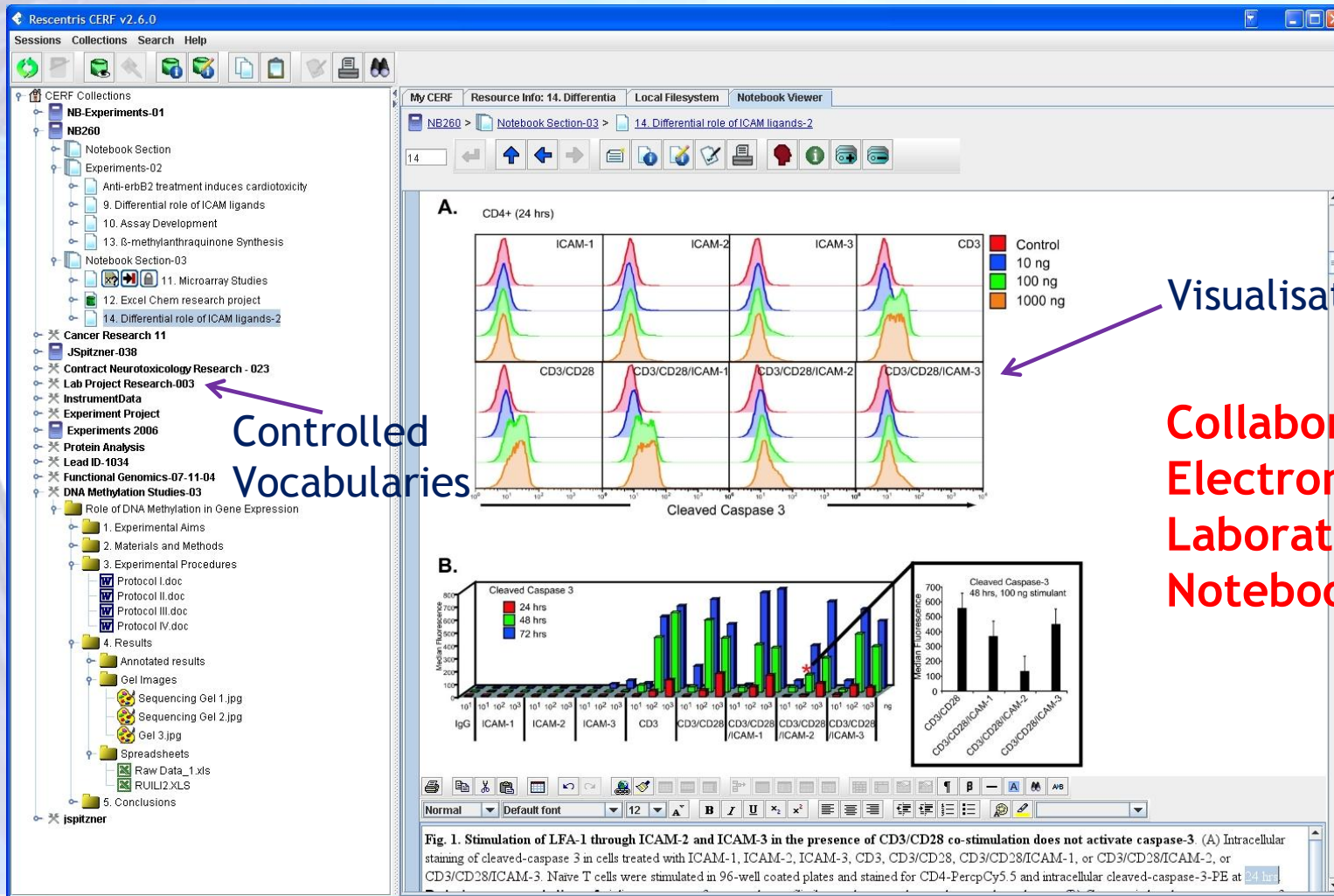


DOCK's activity hits prediction: compound-IDs 005 and 006

Synergy Drug Design Collaboration Pilot



Recording of Collaborative R&D



Controlled
Vocabularies

Visualisation

Collaborative
Electronic
Laboratory
Notebook (ELN)

Collaboration Results Schema

We have defined a Collaboration Results Matrix and XML Schema for the recording of all predictions, experimental data and metadata

<Compound>

<Id>

Unique ID number for each compound, to be used in the pilot

</Id>

<Alternative Names / IDs>

Here we can give vendor codes, IUPAC names, arbitrary names, etc.

<Alternative Name / ID 1>

<ID type>

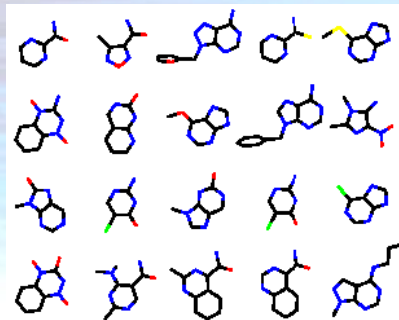
DrugBank ID

</ID type> ETC

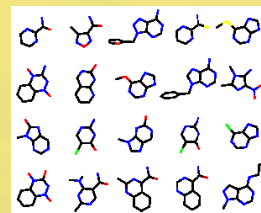
Predictive Toxicity Dashboard

Compound-ID	PREDICTIVE								EXPERIMENTAL				Comments
	Toxicity of Compound				Selection				AFFINITY				
	Predicted Toxicity (Human Adverse Drug Events)	Predicted Toxicity (in vitro, in vivo)	Predicted Toxicity (in silico)	Toxicity Prediction Traffic Light	Weight of Evidence and Quality Analysis	Quality Traffic Light			Low Content Toxicity Assay	Toxicity Assay Traffic Light		High Content Toxicity Assay	
001	1	0	1	“G”	1	“G”	1	“G”	0	“R”	n.d.	“R”	
002	0	1	0	“R”	n.d.	“R”	n.d.	“R”	n.d.	“R”	n.d.	“R”	
003	1	1	0	“Y”	1	“G”	1	“G”	1	“G”	1	“G”	
004	1	1	1	“G”	1	“G”	1	“G”	1	“G”	0	“R”	
004	0	1	1	“G”	1	“G”	0	“R”	n.d.	“R”	n.d.	“R”	

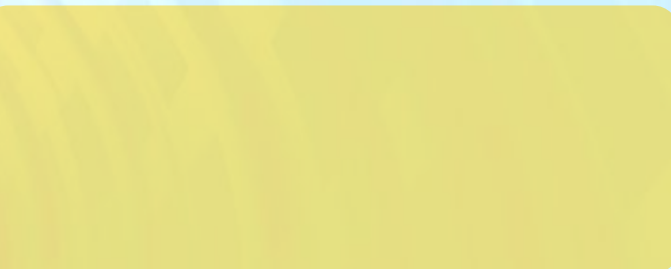
1. A library of compounds is entered to the ELN



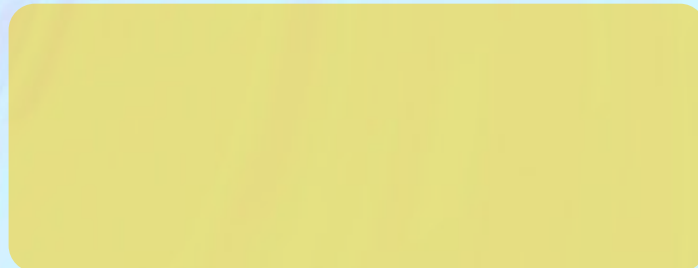
ELN



Synergy

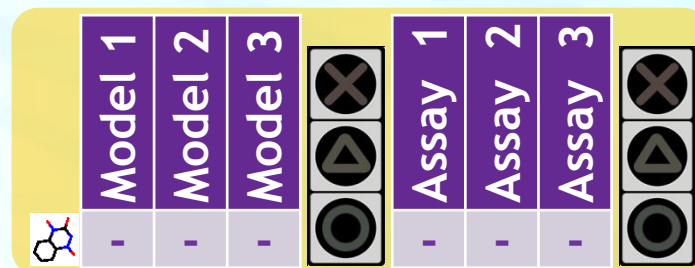


OpenTox

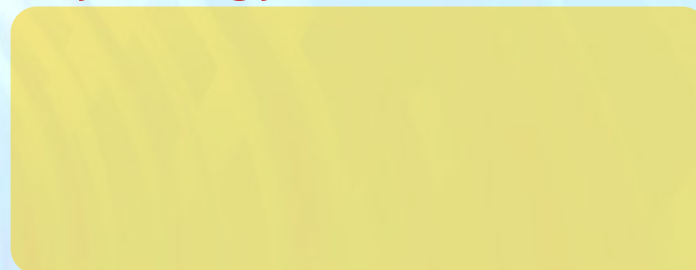


2. Each compound is assigned a data structure in ELN

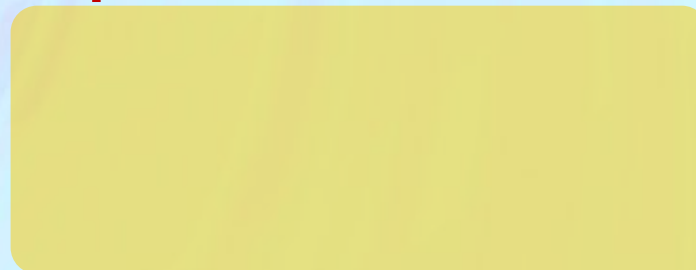
ELN



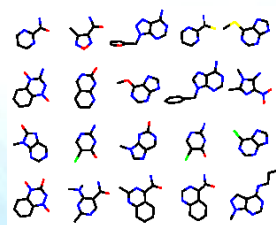
Synergy



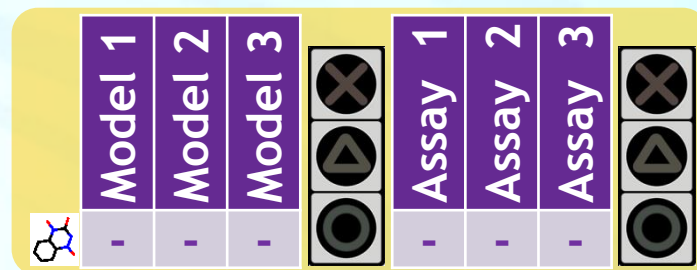
OpenTox



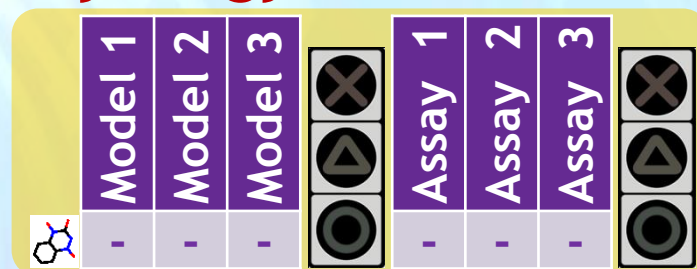
3. ELN passes compounds to OpenTox and SYNERGY



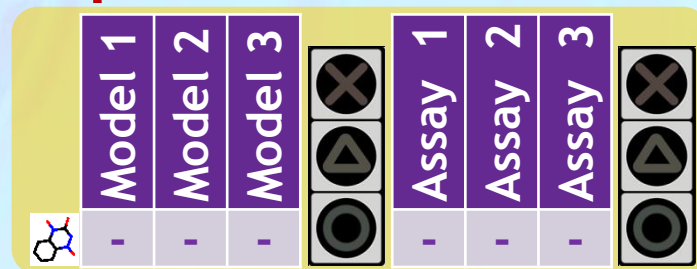
ELN



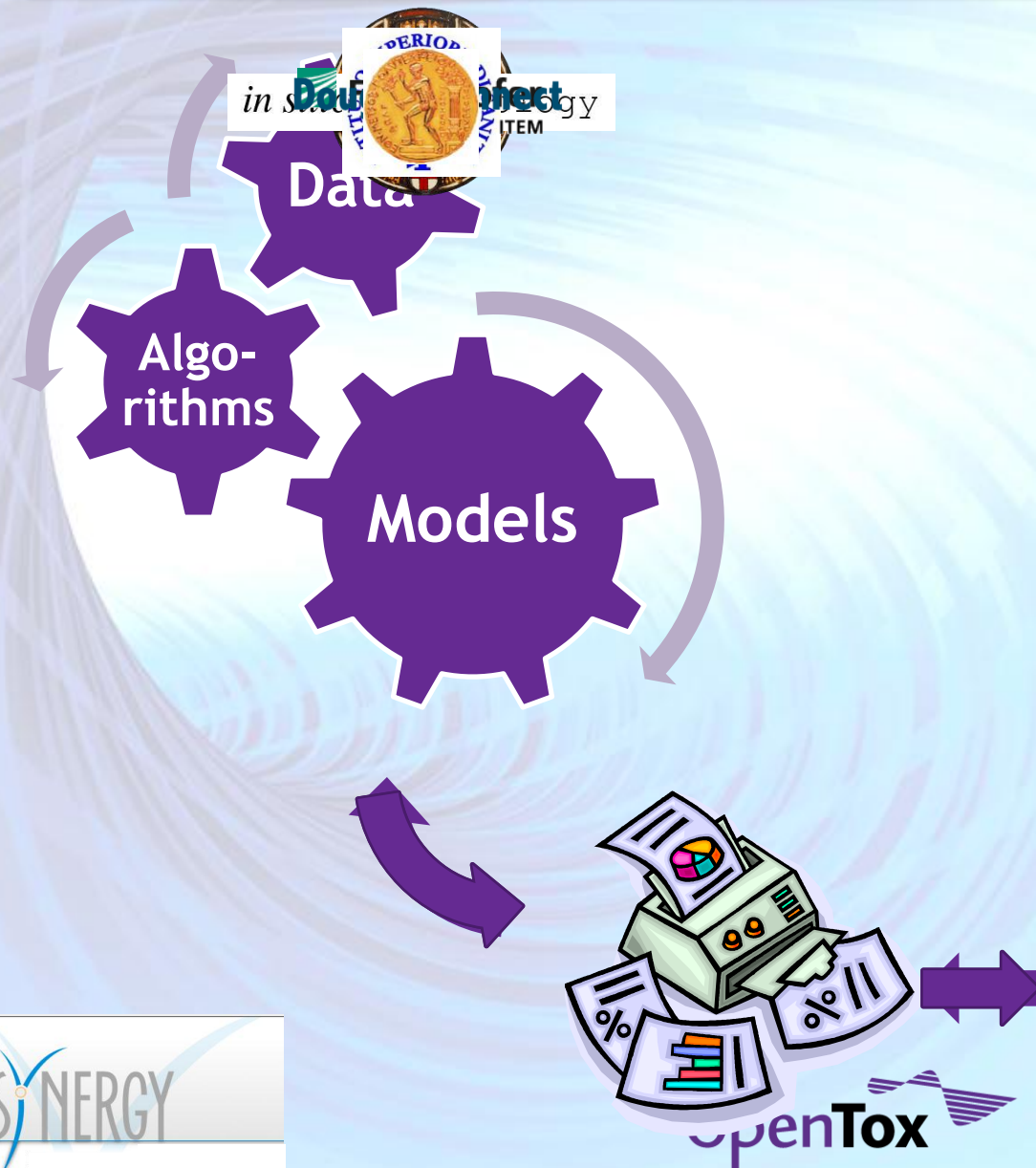
Synergy










OpenTox










4. OpenTox computes toxicity predictions










ELN

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	-	-	-		-	-	-	
	-	-	-		-	-	-	
	-	-	-		-	-	-	

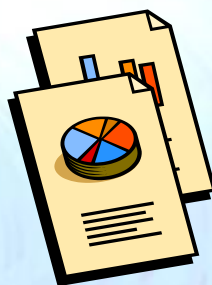
Synergy

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	-	-	-		-	-	-	
	-	-	-		-	-	-	
	-	-	-		-	-	-	








OpenTox

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	1	0	1		-	-	-	
	-	-	-		-	-	-	
	-	-	-		-	-	-	








5. OpenTox sends back a report to ELN










ELN

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	1	0	1		-	-	-	
								
								

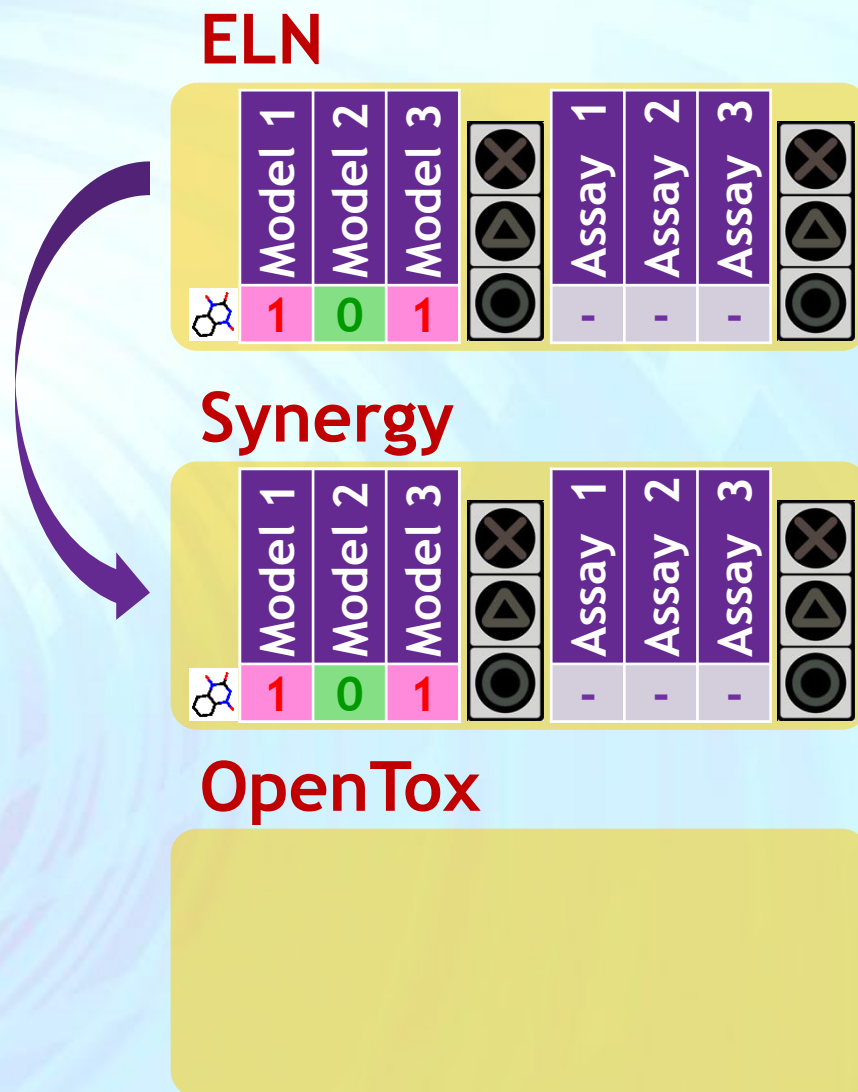
Synergy

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	-	-	-		-	-	-	
								
								

OpenTox

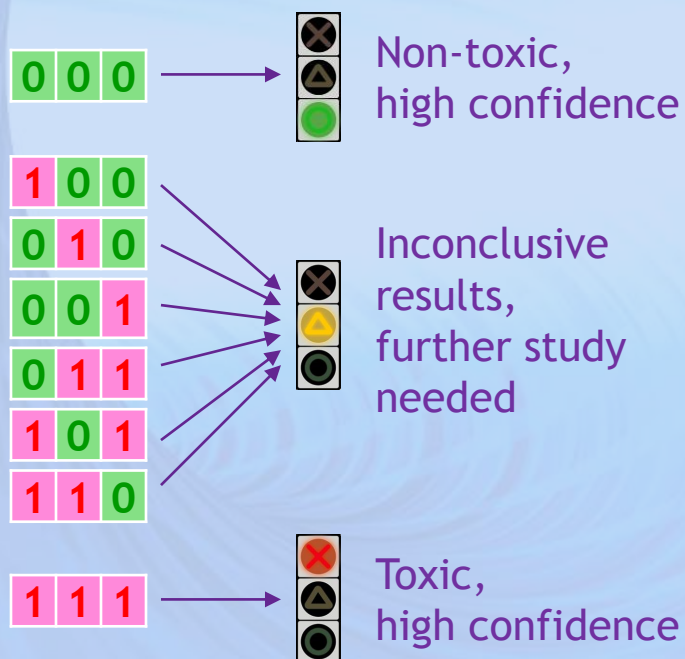
	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	1	0	1		-	-	-	
								
								

6. ELN sends the results to SYNERGY

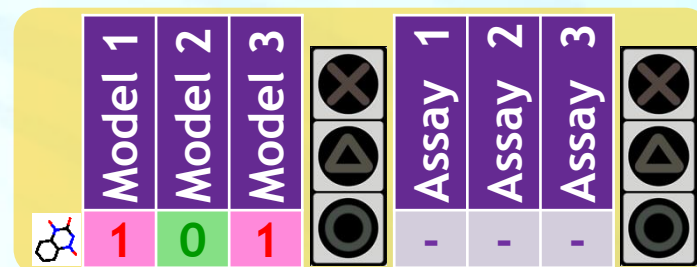


7. SYNERGY applies the Recommendation Rules

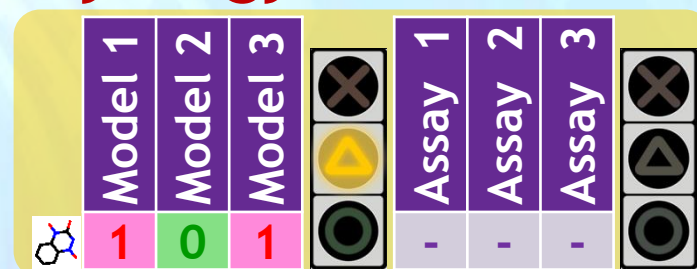
Recommendation Rules:



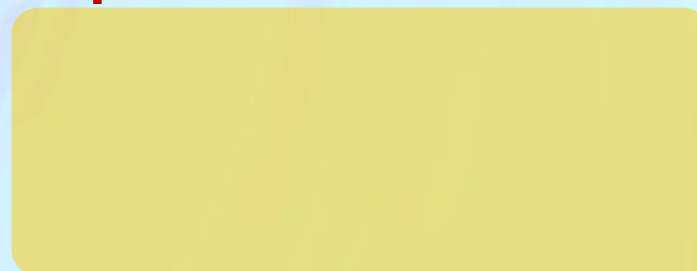
ELN



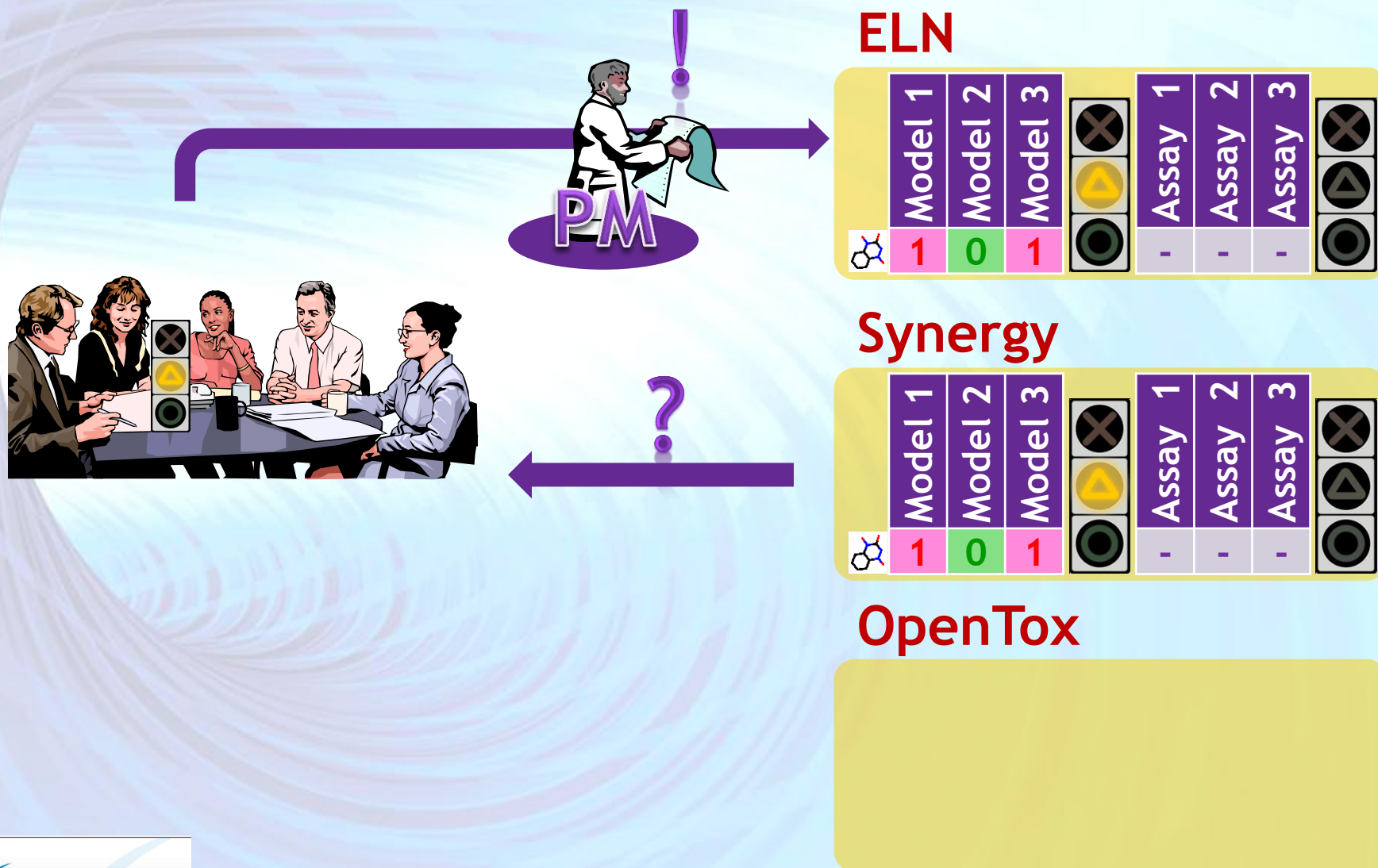
Synergy



OpenTox












8. Inconclusive data → SYNERGY calls a meeting








9. Experimental assays confirm toxicity



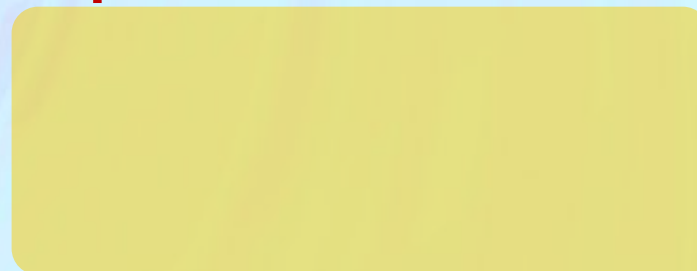
ELN

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	1	0	1		-	1	1	
								

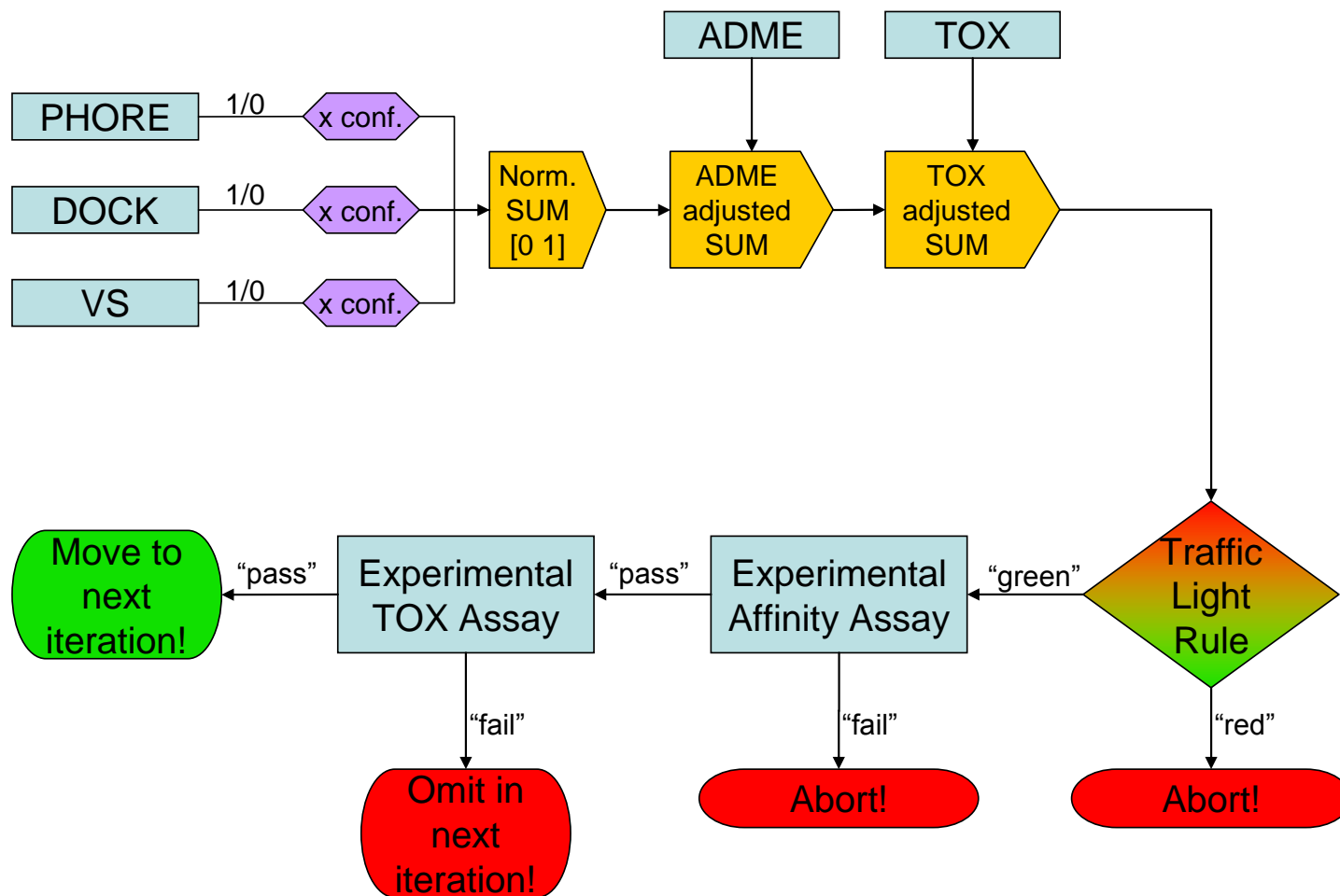
Synergy

	Model 1	Model 2	Model 3		Assay 1	Assay 2	Assay 3	
	1	0	1		-	-	-	
								

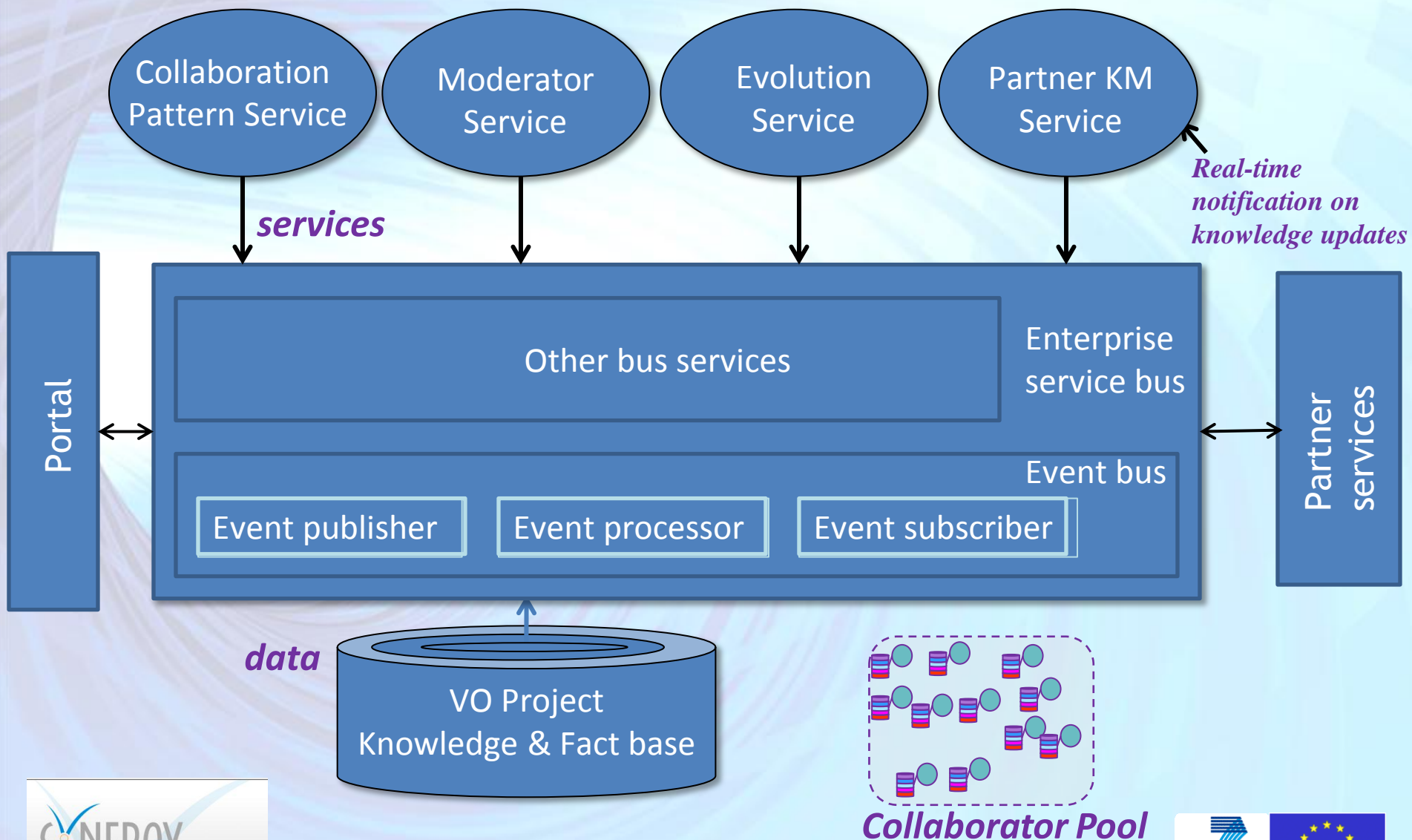
OpenTox



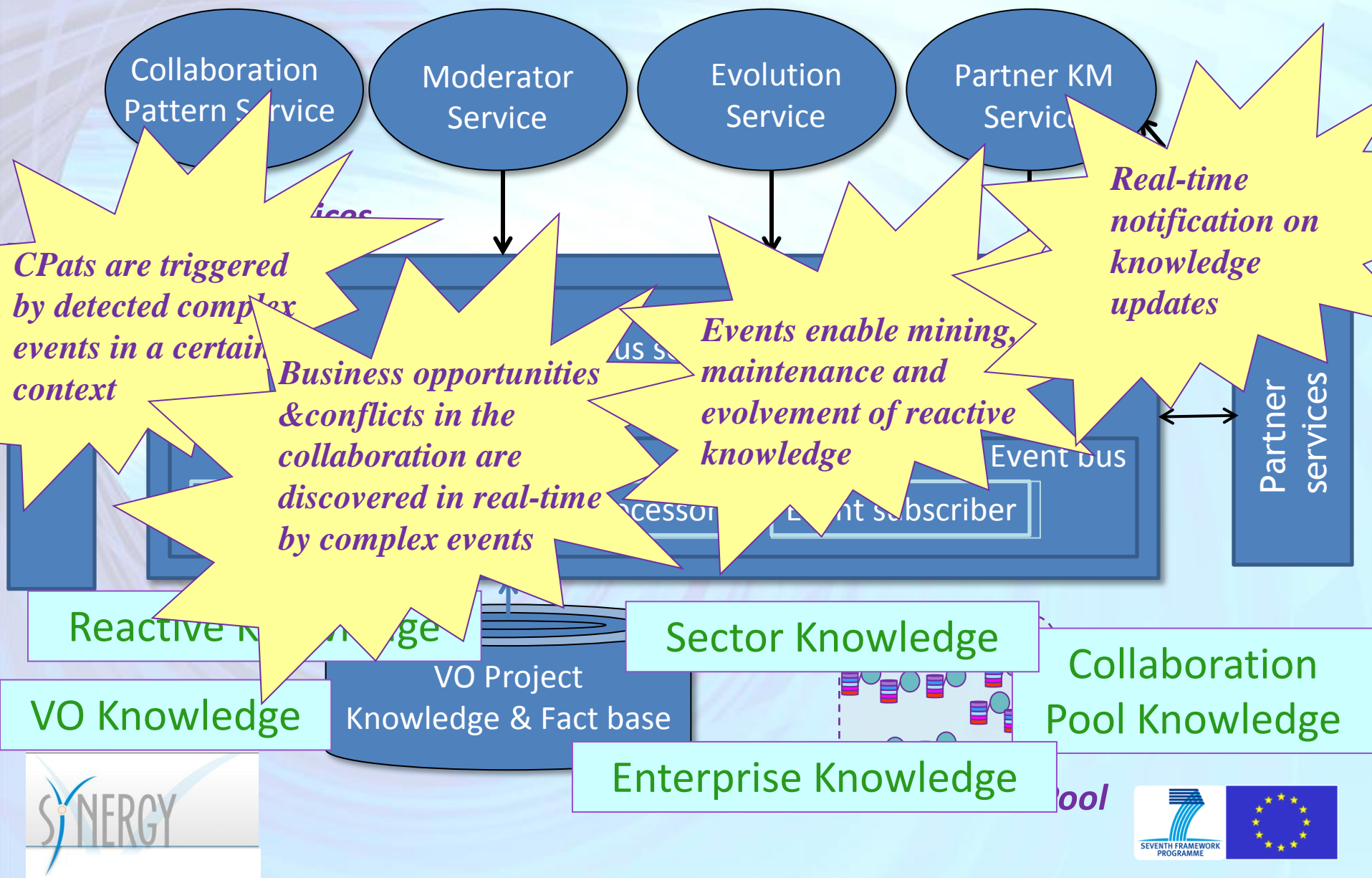
Prediction Workflow



SYNERGY Service support



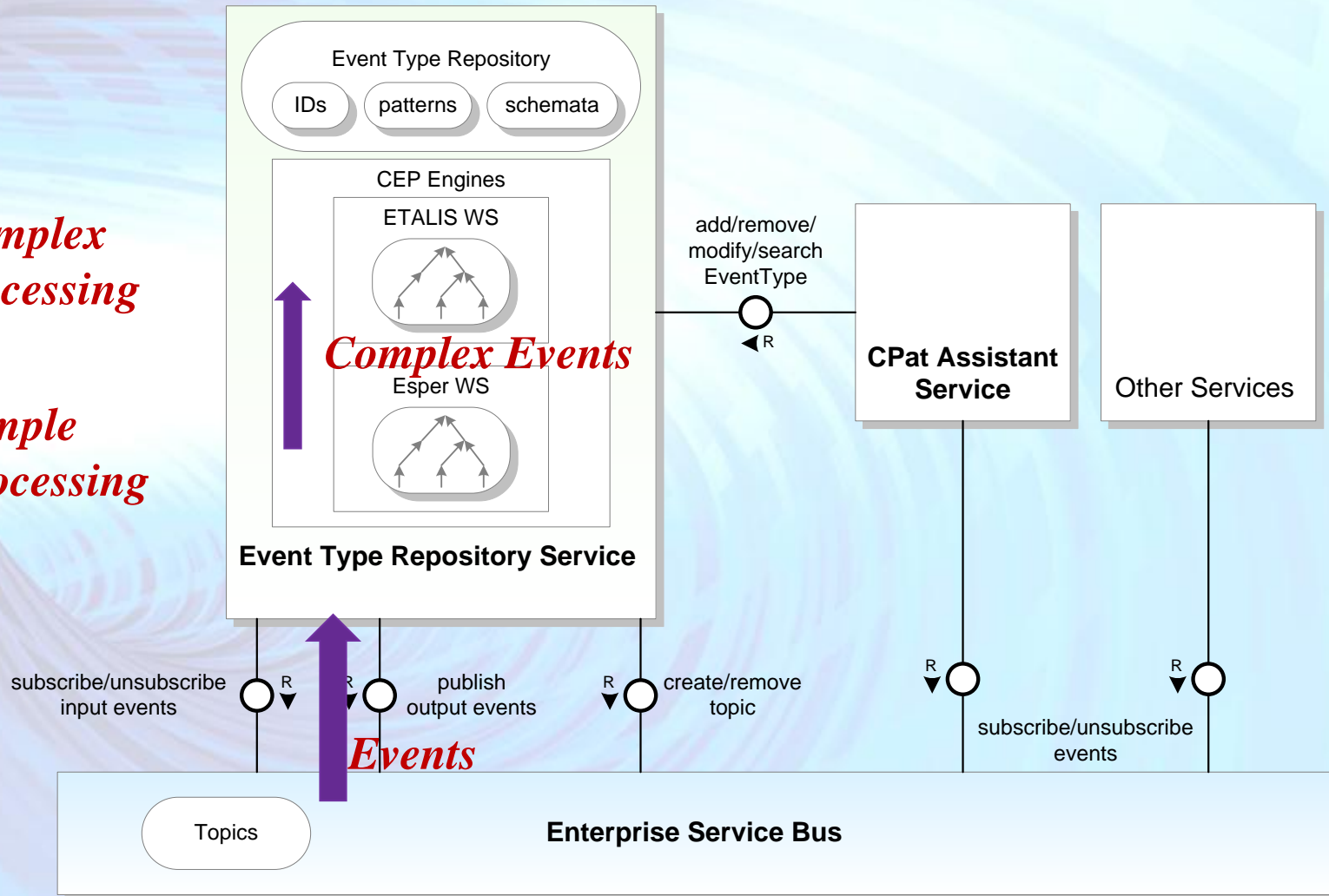
SYNERGY Service support



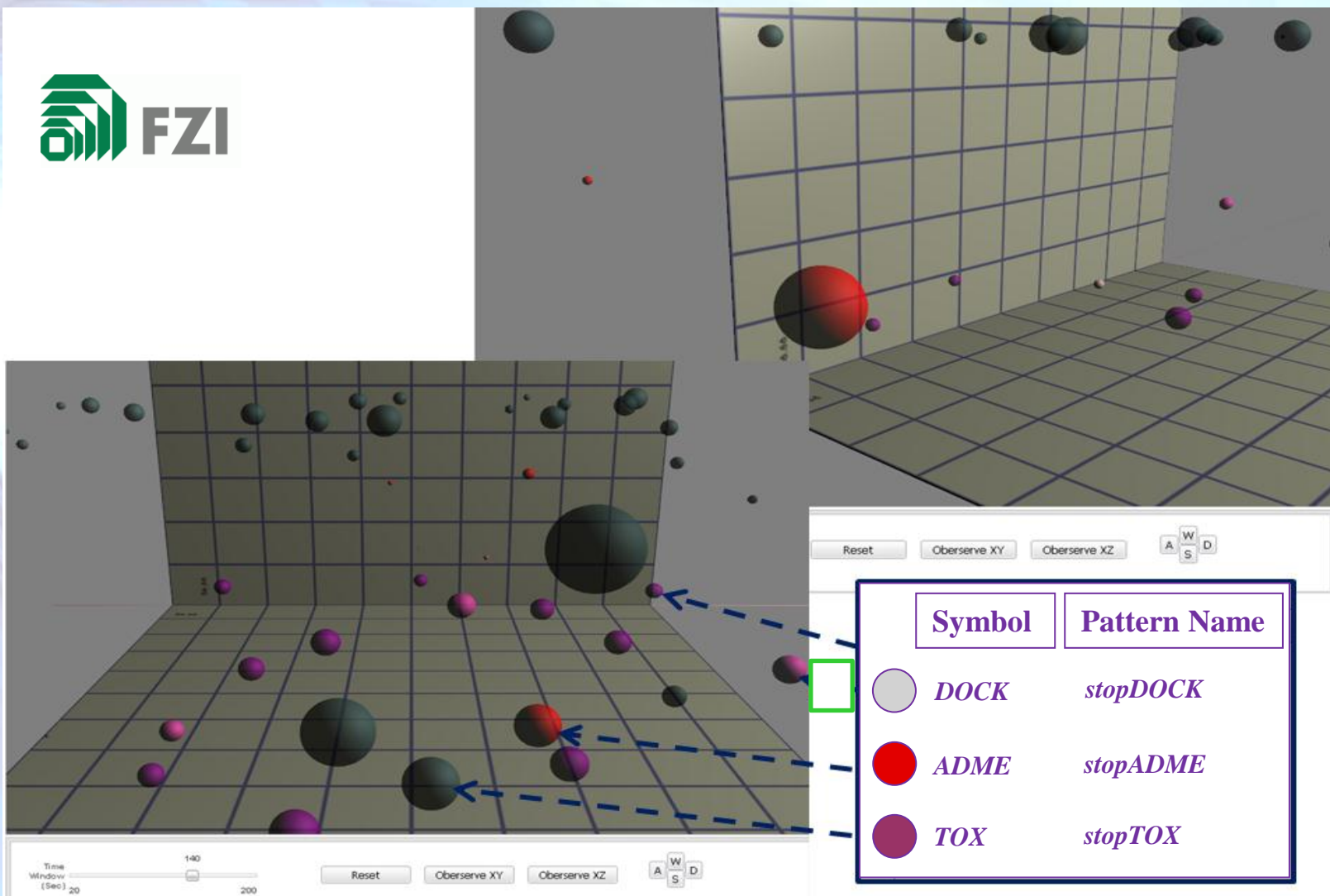
SYNERGY Complex Event Processing

*Complex
processing*

*Simple
processing*

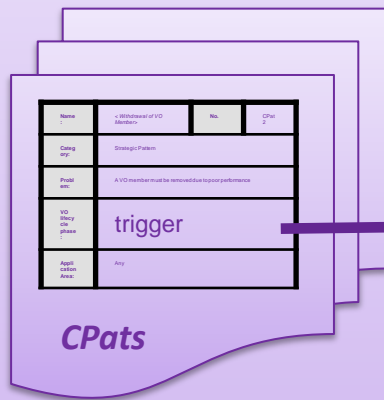


SYNERGY Event Processing of Drug Design Results



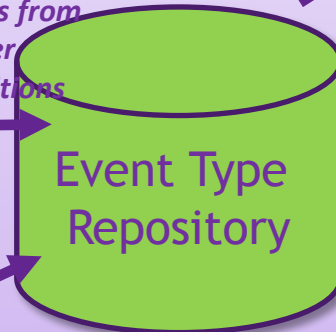
SYNERGY Collaboration Patterns

How are the building blocks of communication (i.e. events) identified?



Where are they represented?

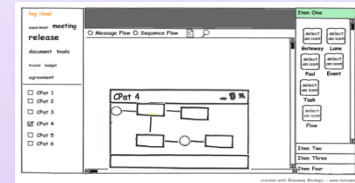
events from trigger definitions



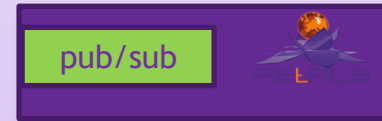
other communication primitives in SYNERGY

1. Identifier
2. Pattern
3. Schema

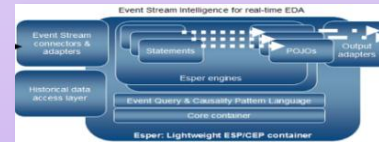
Where are they put to use?



CPat Editor



PEtAls ESB



Esper CEP



ETALIS iCEP

Which are the end consumers?

SYNERGY Services

Facing the Challenges of the Expectations



[Visit with Lions at Mukuni Project, Livingstone, Zambia](#)

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).