

Integrating Predictive Toxicology Applications & Resources

An OpenTox Workshop

30 May, 2010

Potsdam, Germany

Workshop Overview

Barry Hardy
Douglas Connect
OpenTox Project Coordinator

Workshop Overview

1

- Introduction & Overview

2

- Stakeholder Perspectives

3

- Infrastructure Requirements & Interoperability
- OpenTox Framework Design

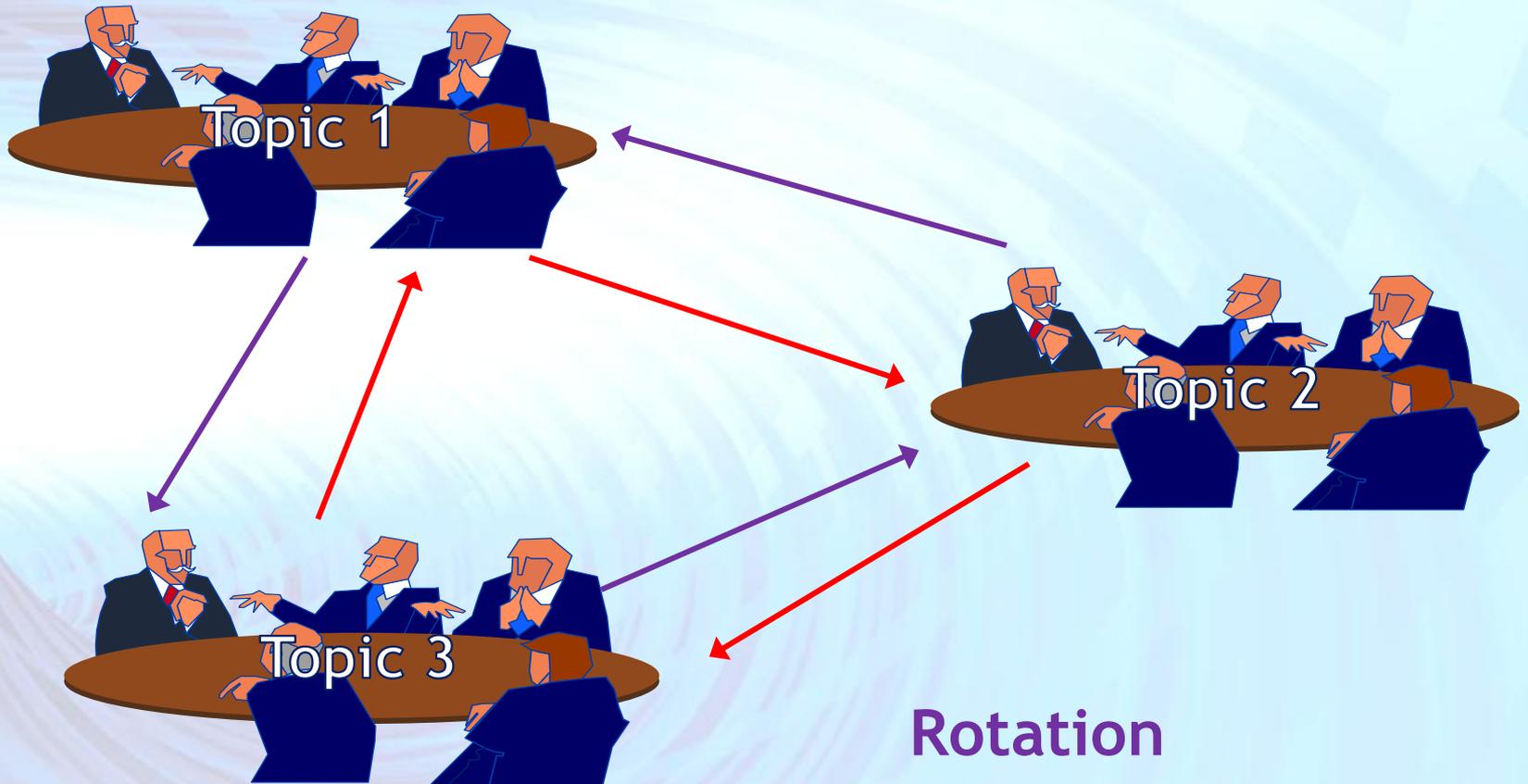
4

- OpenTox Integrating Application Demonstrations
- Extensibility, Sustainability & Future Directions

5

- Knowledge Café Discussions
- Concluding Discussion & Actions

Knowledge Café Discussion Format



Development of Strategies for Interoperable Resources & Applications in Predictive Toxicology



Eliminate
traditional
circus acts,
animals



Create
theatrical
themes,
storylines,
new acts

**Blue
Ocean**

Reduce
dangerous acts,
traditional
humour,
transport costs



Raise
tent standards,
artistic
sophistication,
ticket prices!



Based on Blue Ocean Strategy,
Kim & Mauborne 2006

Perspectives

Robert Kavlock (EPA, US)

Carl Westmoreland (Unilever, UK)

Emilio Benfenati (Mario Negri Institute, Italy)

Egon Willighagen (Uppsala University, Sweden)

Jeffrey Wiseman (Pharmatropé, US)

Michael Schwarz (University of Tuebingen, Germany)

Requirements in Predictive Toxicology Infrastructure

Barry Hardy
Douglas Connect
OpenTox Project Coordinator

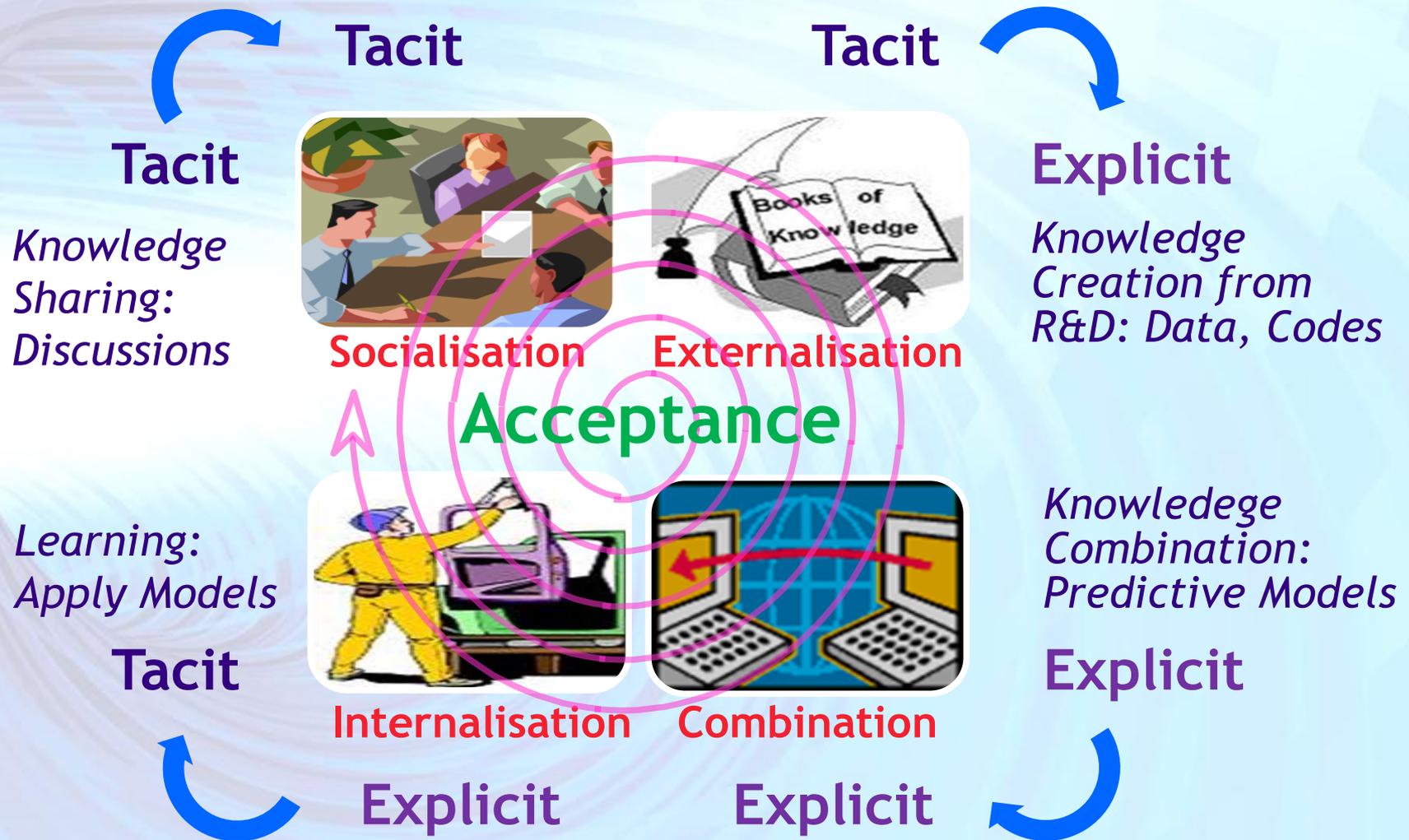
Need for communications in the community overcoming different languages and vocabularies



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

[From Conservation Project Trip in Caprivi Delta](#)

SECI Model for Knowledge Management



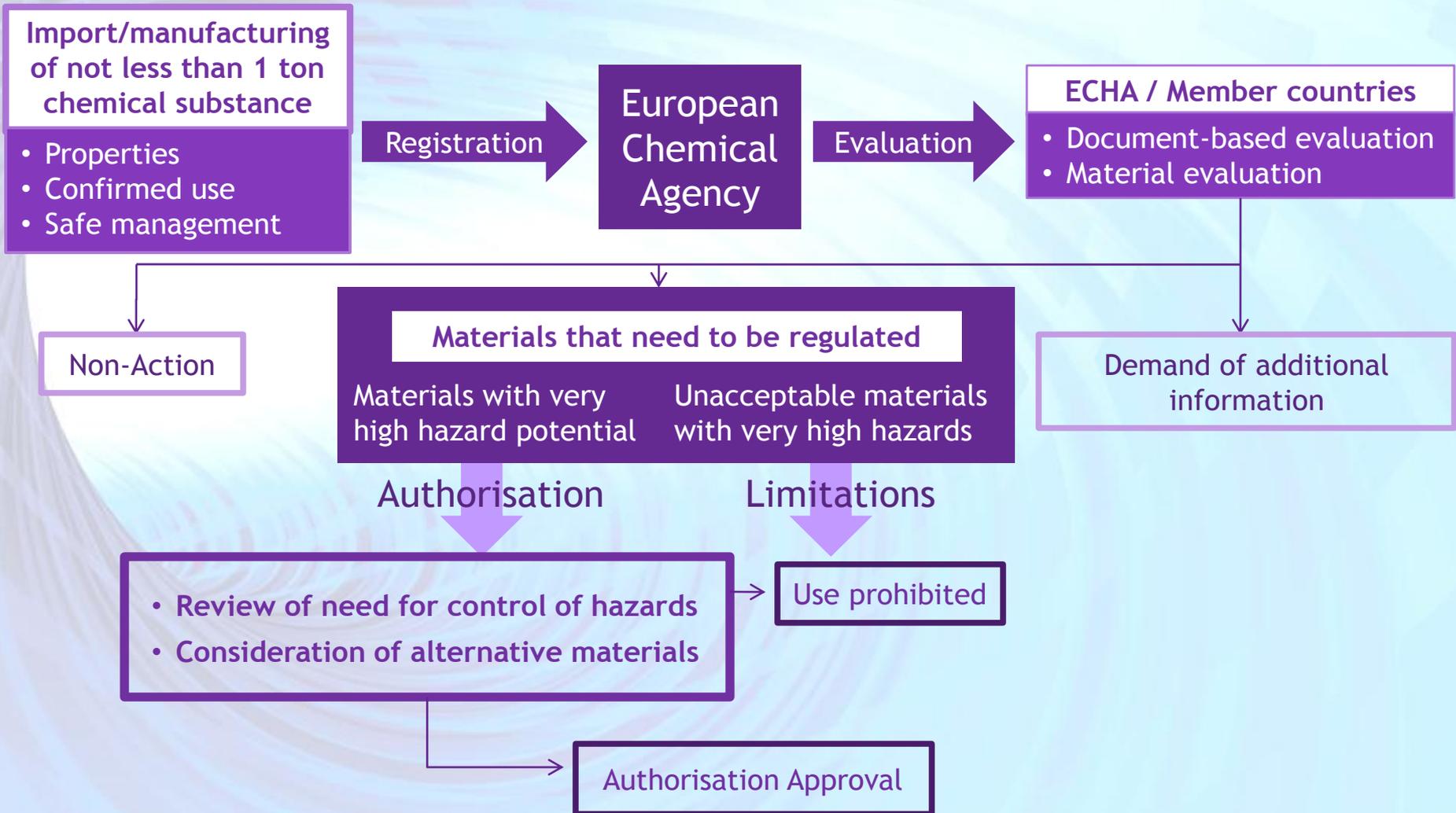
REACH



REACH

(enRegistrement, Evaluation et
Autorisation des substances CHimiques)

REACH Registration



What is our situation today when we say ...

**Let's build an
integrated predictive
toxicology application!**

**Let's collaborate on an
analysis!**



Faced with such an integration challenge ...



... we tend to look away from some of the problems ...

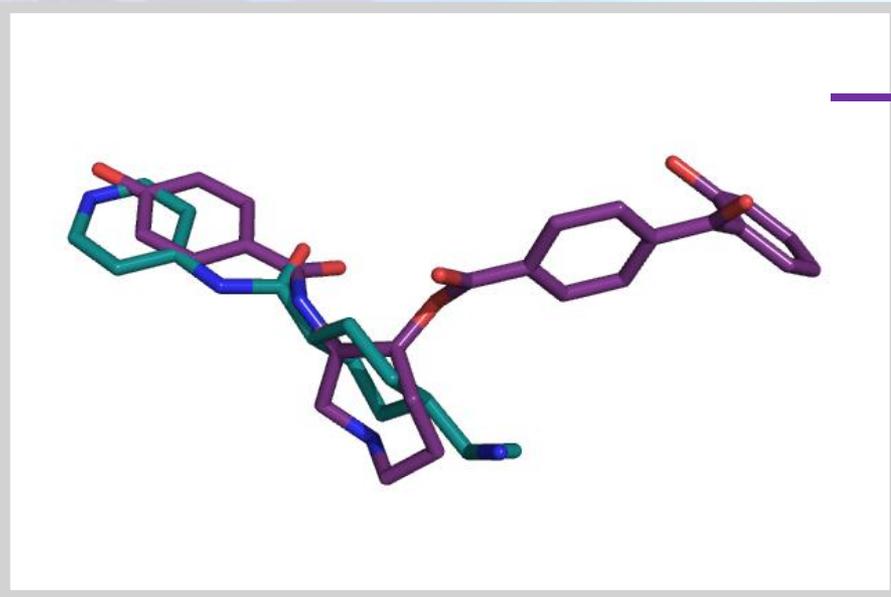
Faced with such an integration challenge ...



... we tend to look away from some of the problems and from the need to collaborate more closely.

Collaborative Predictive Toxicology Challenge

Input Structure

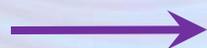


Out - Toxic or Not?

- LD50
- Liver Toxicity
- Secondary Metabolites
- Bioavailability
- Mutagenicity
- Carcogenicity
- Reproductive Toxicology
- Skin Irritation
- Aqua Toxicity
- Combined predictions for arrays of multiple end points



Driver



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

Accelerating Knowledge Flows in Predictive Toxicology

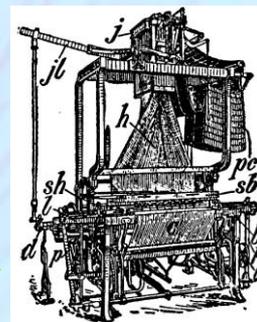
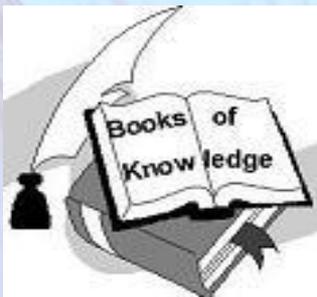
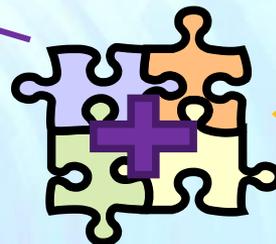
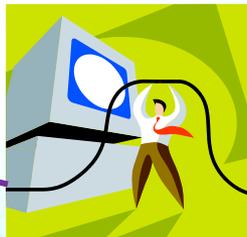
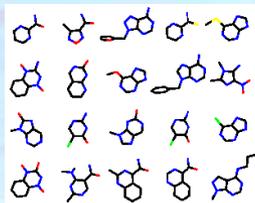
ToxPredict
guest Help Admin

Step 1: Search (find structure)
Step 2: Verify (verify structure)
Step 3: Models (model generation)
Step 4: Evaluate (evaluate model)

Model

Endpoint	Descriptors	Training dataset	Algorithm
MolecularWeight	-	-	MolecularWeight
CP5A descriptor	-	-	CP5A descriptor
CP5A index	-	-	CP5A index
Predictive QSAR Model generated by the algorithm mlr	YES	Multiple Linear Regression Training Algorithm	Multiple Linear Regression Training Algorithm
Predictive QSAR Model generated by the algorithm mlr	YES	Multiple Linear Regression Training Algorithm	Multiple Linear Regression Training Algorithm
Quantile model created with TUM's MLRegression model learning web service.	-	-	http://opentox.informatik.tu-muenchen.de/080/Opentox-dev/algorithm/mlrregression
Acute toxicity to fish (lethality)	YES	-	-
ToxTree	-	-	-

Toxicity Predictions



Compelling Needs of Users

Integrated Testing

in silico

in vitro

TTC

Read
Across

Category
Formation

REACH Reporting
(QPRF, QMRF)

Applicability
Domain

Validation

Human
Data

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

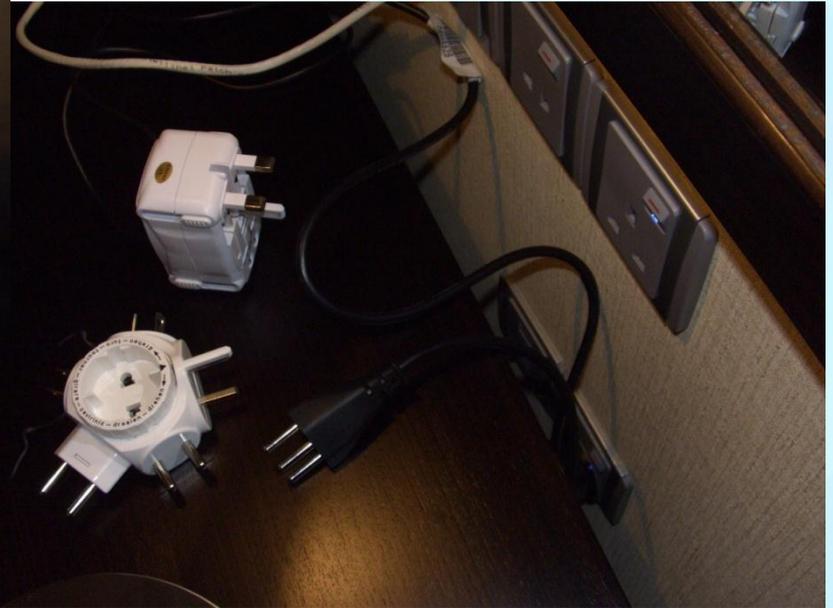
Categories

Systems
Biology

Challenges to Integrated Resources & Applications

- Database silos
- Missing information
- Varying quality
- Hard to integrate data
- Hard to integrate models
- No common framework
- Lack of standards
- Lack of validation
- Complex subject
- Application difficult
- Lack of transparency
- Interdisciplinary collaboration

Absence of Interoperability creates Problems



Adaptor Challenge in Jeddah, 2008

Interacting Components create Solutions



Adaptor Solution in Jeddah, 2008

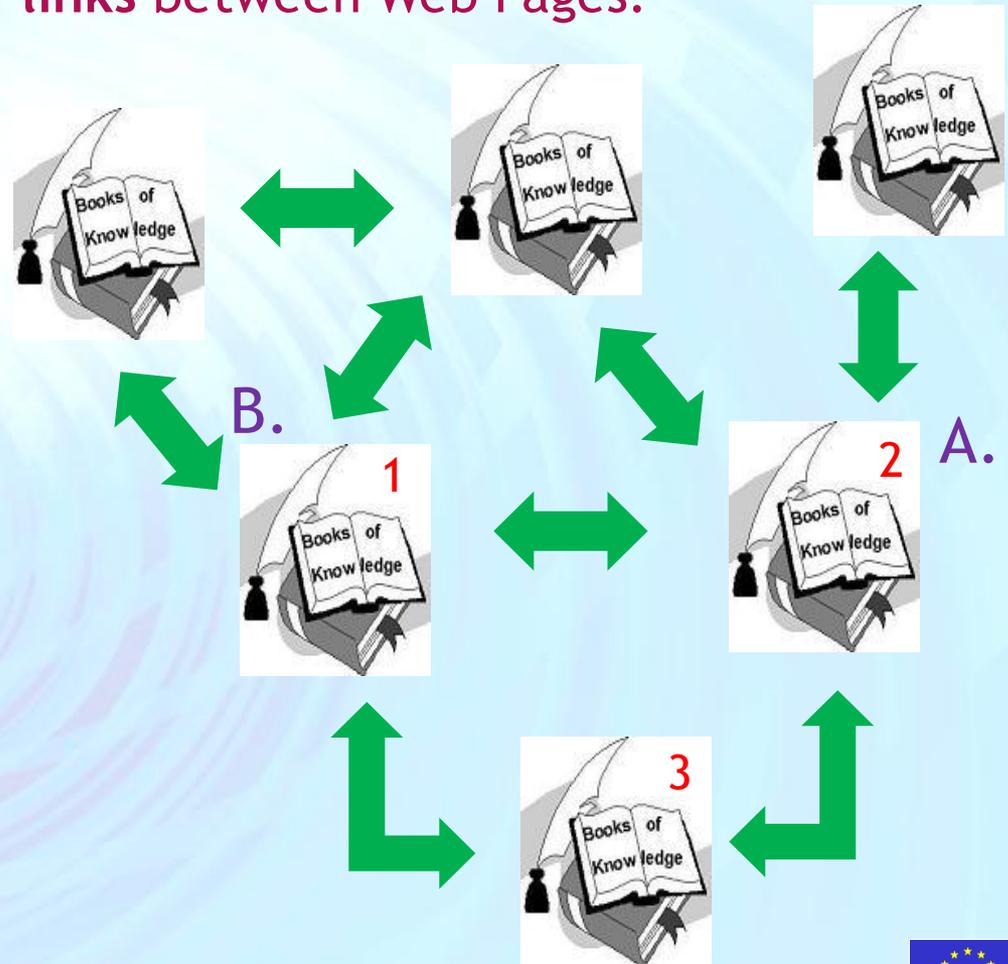
Value is in Linking

The key idea of Google's founders in creating their search engine:
There is useful knowledge in the **links** between Web Pages.

Page Ranking

A page is ranked higher in a search if:

- A. it has more connections to it than other pages
- B. the pages connecting to it have higher ranking themselves



Solution created by Linked Open Data, Web Applications and Crowdsourcing



Haiti Earthquake Crisis Response (2010)

wiki.openstreetmap.org

Interoperability & Vocabulary

Dogs

- Collie
- Labrador



Cats

- Siamese
- Persian



Birds

- Sparrow
- Owl



Interoperability & Vocabulary

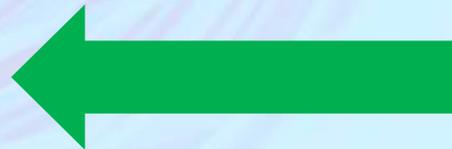


Interoperability & Ontology

Org A



Org B



OpenTox Approach

Framework

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

Diverse Access

- Toxicologists
- Computational Scientists
- Interfaces for new algorithm development & integration

Interoperability

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

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



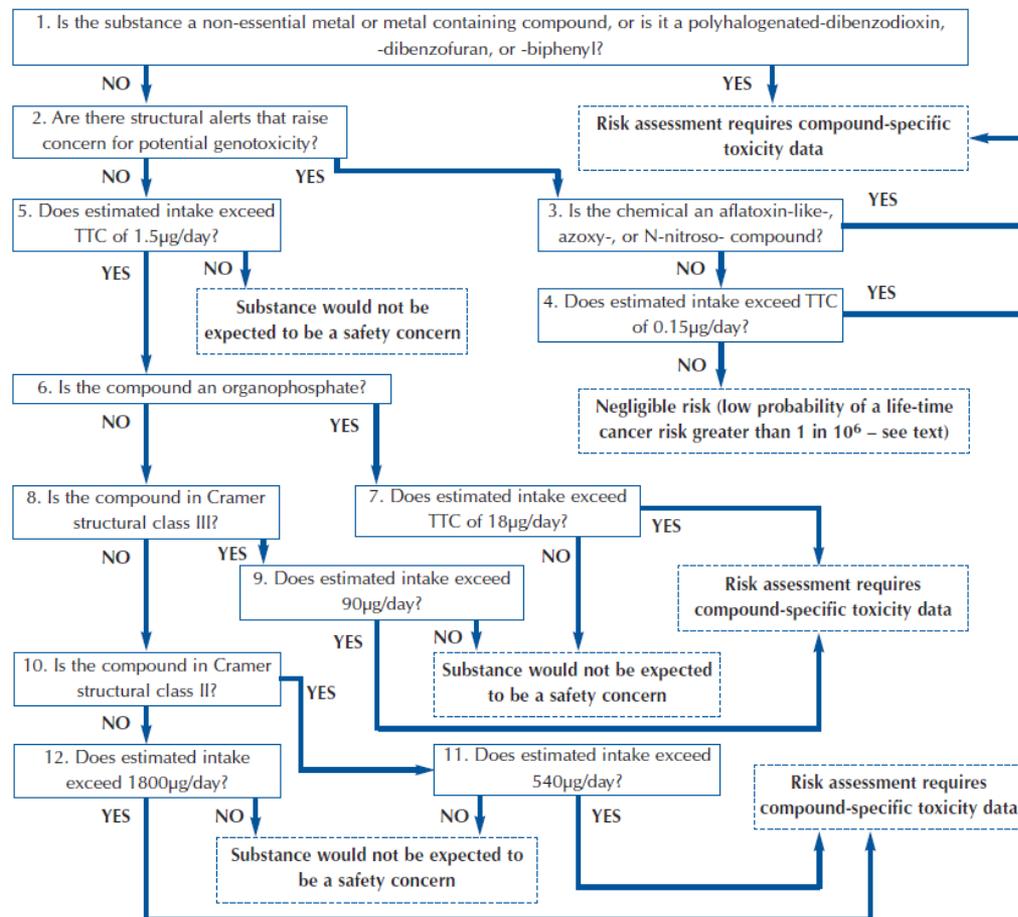
www.opentox.org/opentoxipedia



OpenTox Use Case Example

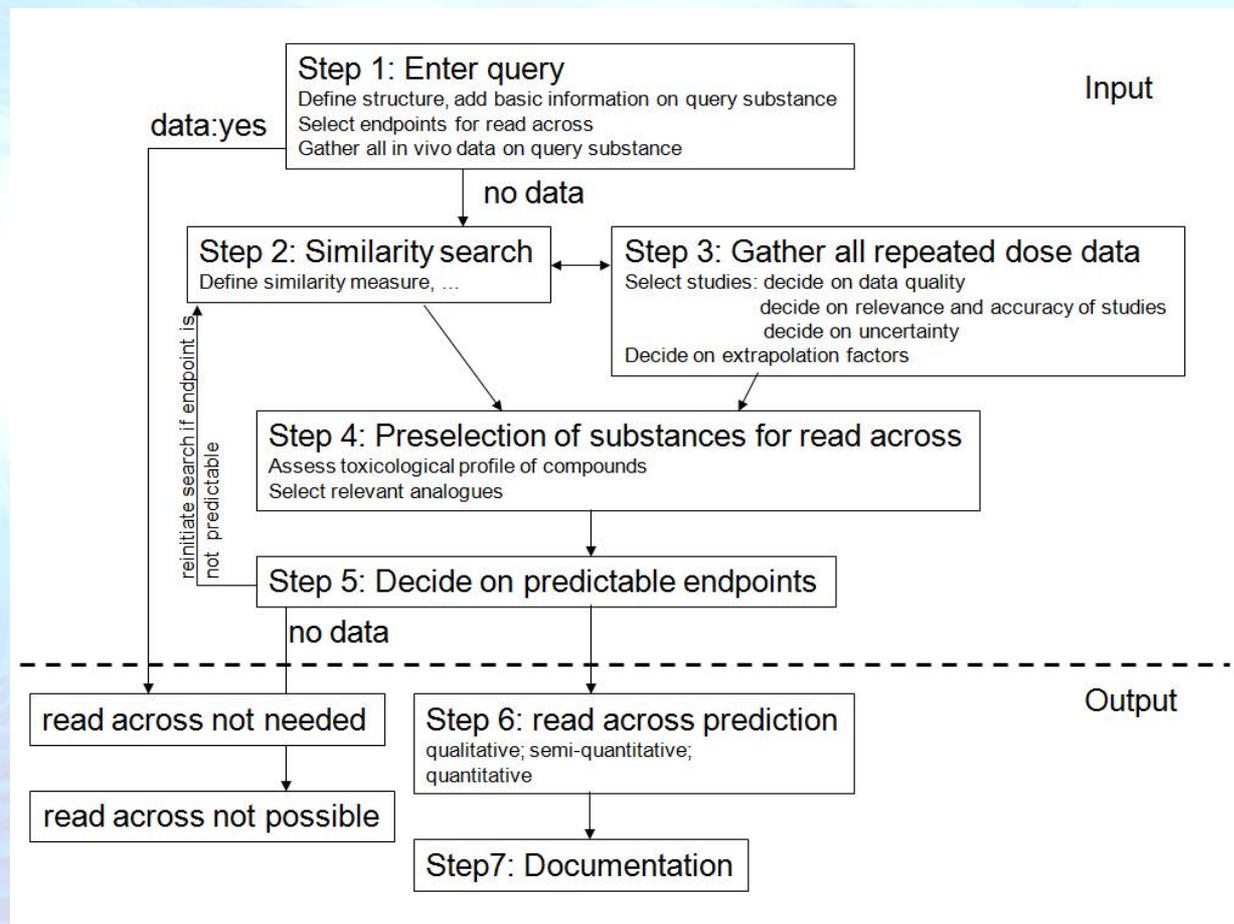
Implement Threshold of Toxicological Concern (TTC) using ILSI decision tree guidelines

Decision tree proposed by ILSI Europe to decide whether substances can be assessed by the TTC approach (From Kroes *et al.*, *Food and Chemical Toxicology* 42, p76, 2004)



OpenTox Use Case Example

Support Read Across for Repeated Dose Toxicity



Step 1: Search
Select structure(s)

Step 2: Verify structure
Verify structure

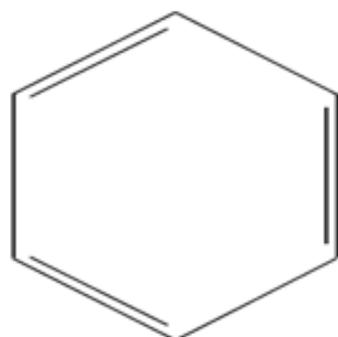
Step 3: Models
Select prediction models

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

	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 prediction of toxicological mechanisms and the recording of opinions and analysis in reports

The OpenTox Framework Design

Stefan Kramer

Technical University of Munich

OpenTox Approach

Framework

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

Diverse Access

- Toxicologists
- Computational Scientists
- Interfaces for new algorithm development & integration

Interoperability

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

Strategic Context/Goals

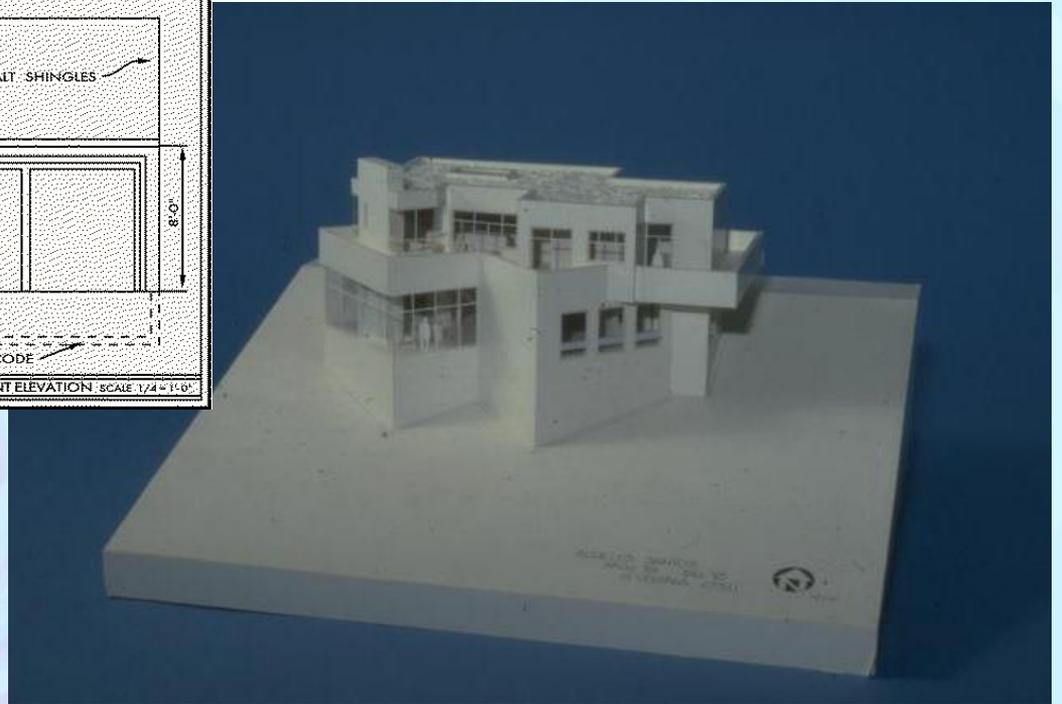
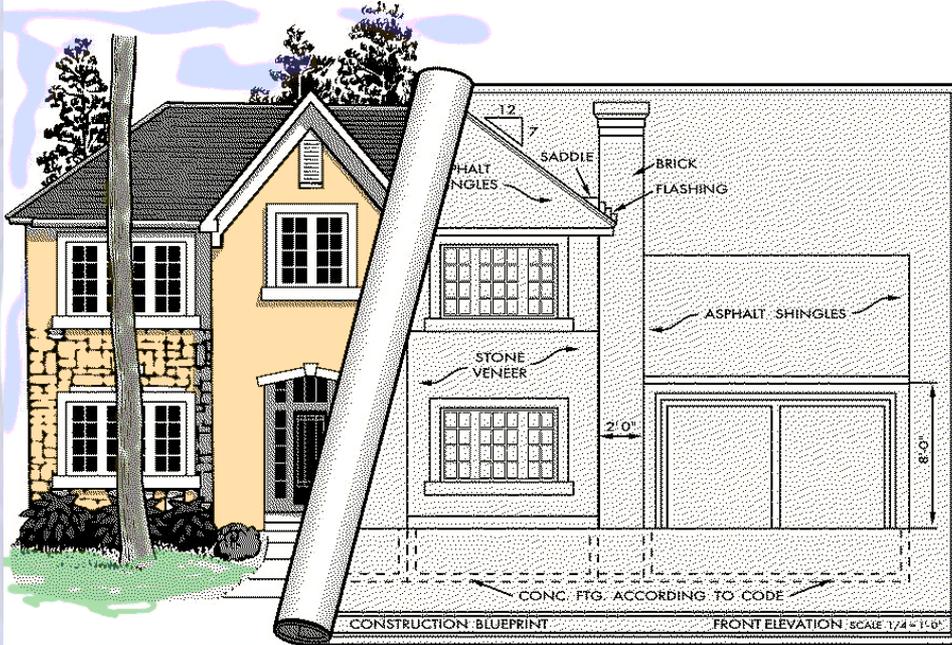
- REACH: possible reduction of test animals by using existing experimental data in conjunction with QSAR
- Also practical needs: **reporting** and form filling
- By the OECD principles, a number of requirements to a framework like OpenTox arise

	OECD Principle	OpenTox addresses 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 prediction of toxicological mechanisms and the recording of opinions and analysis in reports

Consequences for Requirements on OpenTox

User Requirements		Software Requirements
Unambiguous data	⇒	<i>formal way of representing information about data</i>
Unambiguous access	⇒	<i>well-defined interfaces</i>
Transparency of computational tools	⇒	<i>formal way of representing information about methods, well-defined interfaces</i>
Variety of user groups	⇒	<i>simplicity and modularity of design</i>
Need to integrate various resources (e.g., databases, prediction methods, models, ...) to make meaningful predictions	⇒	<i>distributed architecture, interoperability</i>
Need to integrate biological information	⇒	<i>again, modularity of design, extensibility</i>

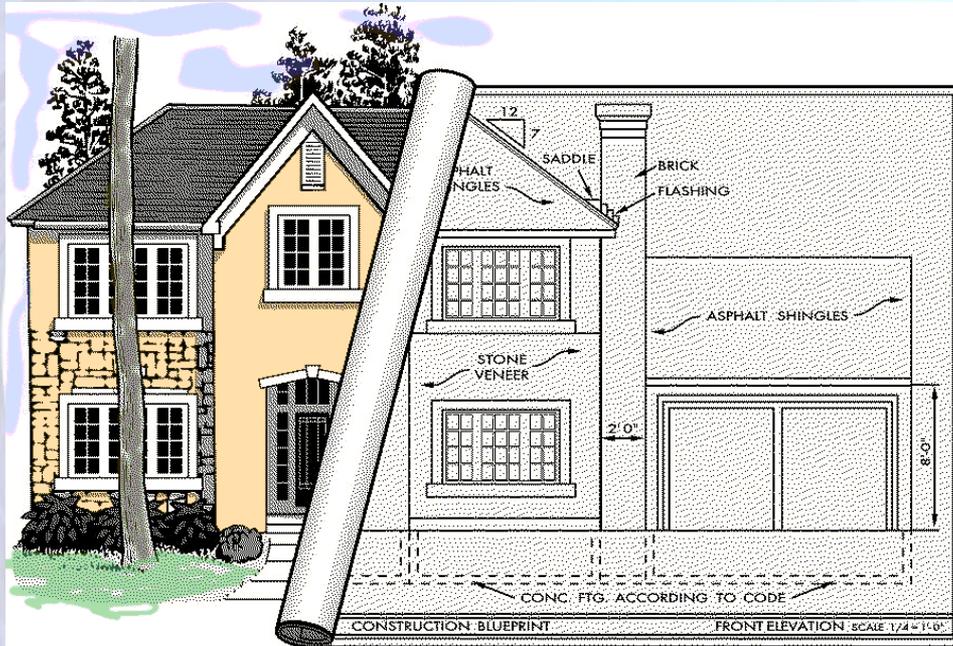
Software Architecture



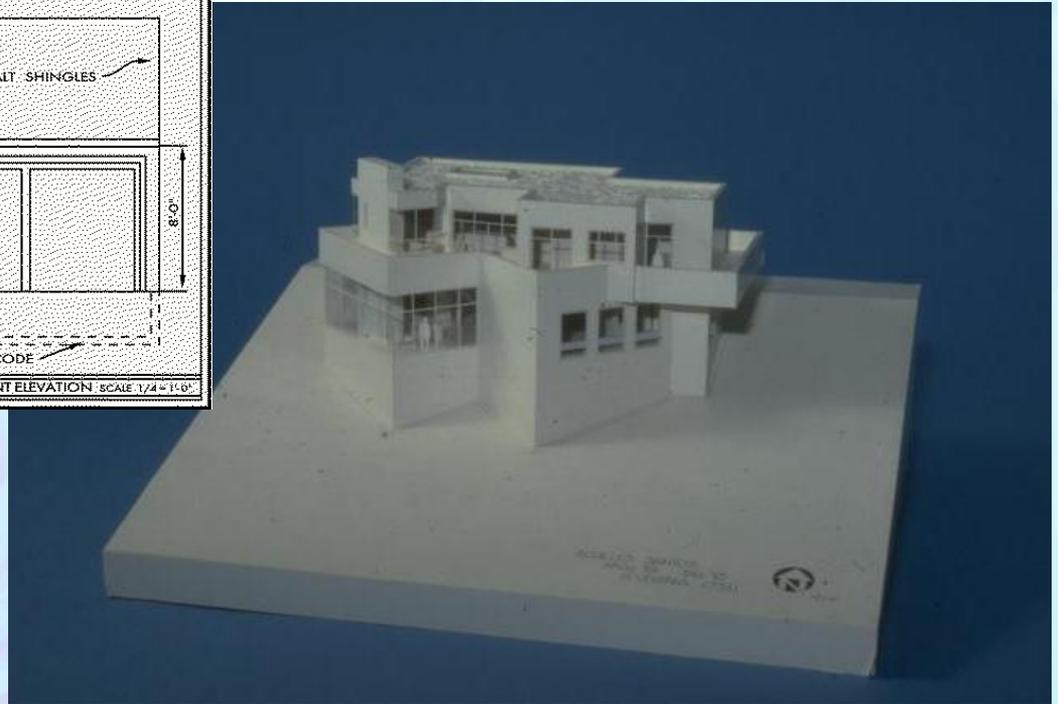
Software Architecture



Software Architecture



To solve a wide range of modelling tasks: needs buildings blocks (modularity) and well-defined ways of combining them



Technological choices...

Representational State Transfer (REST): What and Why?

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 systems**
- (More or less) language independent/installation-free

Ontologies: What and Why?

What?

- **Formal, shared conceptualization** of a **domain**

Why?

- Distributed services **need** to be able to “talk to each other”, i.e. have a **common understanding** of endpoints, any type of property, methods, etc.



Ontologies

- Standards: **OWL** as representation language and **SPARQL** as query language
- There are many ongoing biological ontology projects
- Our strategy: use existing work and standards wherever possible
- However, there are new ontology needs for OpenTox applications, e.g. for algorithms, toxicological endpoints

OpenTox
Ontology Working Group

The screenshot shows a 'CLASS BROWSER' window for the project 'AlgorithmTypes'. The window displays a 'Class Hierarchy' tree. The root is 'owl:Thing (14)'. Underneath, there is a category 'ota:AlgorithmType' which is expanded to show several sub-classes. The 'ota:Learning' sub-class is further expanded, and 'ota:Classification (4)' is highlighted. Other sub-classes include 'ota:DescriptorCalculation', 'ota:MSDMTox', 'ota:Preprocessing', and 'ota:Utility'. Each class is represented by a yellow circle icon and a text label with a count in parentheses.

```
graph TD
    owlThing["owl:Thing (14)"]
    otaAlgorithmType["ota:AlgorithmType"]
    otaDescriptorCalculation["ota:DescriptorCalculation"]
    otaMSDMTox["ota:MSDMTox"]
    otaPreprocessing["ota:Preprocessing"]
    otaUtility["ota:Utility"]
    otaClassification["ota:Classification (4)"]
    otaEagerLearning["ota:EagerLearning (4)"]
    otaLazyLearning["ota:LazyLearning (4)"]
    otaRegression["ota:Regression (4)"]
    otaMultipleTargets["ota:MultipleTargets (4)"]
    otaRules["ota:Rules"]
    otaSingleTarget["ota:SingleTarget (4)"]
    otaDataCleanup["ota:DataCleanup"]
    otaDiscretization["ota:Discretization (2)"]
    otaFeatureSelection["ota:FeatureSelection (2)"]
    otaNormalization["ota:Normalization"]
    otaSemiSupervised["ota:SemiSupervised"]
    otaSupervised["ota:Supervised (11)"]
    otaUnsupervised["ota:Unsupervised (3)"]
    otaGeneration3D["ota:Generation3D"]
    otaSimilarityDistanceCalculation["ota:SimilarityDistanceCalculation"]
    otaVisualisation["ota:Visualisation"]

    owlThing --- otaAlgorithmType
    owlThing --- otaDescriptorCalculation
    owlThing --- otaMSDMTox
    owlThing --- otaPreprocessing
    owlThing --- otaUtility
    otaAlgorithmType --- otaPatternMining["ota:PatternMining (2)"]
    otaAlgorithmType --- otaPharmacophoreGeneration["ota:PharmacophoreGeneration"]
    otaAlgorithmType --- otaPhysicoChemical["ota:PhysicoChemical"]
    otaAlgorithmType --- otaQuantumChemical["ota:QuantumChemical"]
    otaAlgorithmType --- otaSimilarityDistance["ota:SimilarityDistance"]
    otaAlgorithmType --- otaTopological["ota:Topological"]
    otaMSDMTox --- otaClustering["ota:Clustering"]
    otaMSDMTox --- otaLearning["ota:Learning"]
    otaLearning --- otaClassification
    otaLearning --- otaEagerLearning
    otaLearning --- otaLazyLearning
    otaLearning --- otaRegression
    otaLearning --- otaMultipleTargets
    otaLearning --- otaRules
    otaLearning --- otaSingleTarget
    otaPreprocessing --- otaDataCleanup
    otaPreprocessing --- otaDiscretization
    otaPreprocessing --- otaFeatureSelection
    otaPreprocessing --- otaNormalization
    otaUtility --- otaSemiSupervised
    otaUtility --- otaSupervised
    otaUtility --- otaUnsupervised
    otaUtility --- otaGeneration3D
    otaUtility --- otaSimilarityDistanceCalculation
    otaUtility --- otaVisualisation
```

OpenTox: Databases

Chemical compounds - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://apps.ideaconsult.net:8180/ambit2/query/smarts?type=smiles&search=[*]OC(=O)[%236%3BH1]%3D[%236%3BH1]c1cccc1&t

Chemical compounds

ToxPredict TTC Depiction Datasets Chemical compounds Similarity Substructure Algorithms References Features Templates Models Ontology RDF playground Help

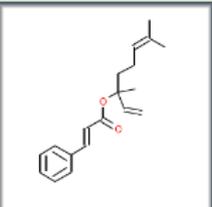
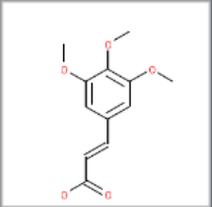
ambit

SMARTS

Keywords

Search for substructure and properties
This site and AMBIT REST services are under development!

Retrieve data Search results SMARTS [*]OC(=O)[#6; Download as Max number of hits:

#	Compound	ECHA REGISTRATION DATE	ECHA CasRN	ECHA EC	ECHA Names	ECHA SYNON Names	ECHA SYNON Names	ECHA SYNON Names	ECHA SYNON Names	ECHA SYNON Names	ECHA SYNON Names
1		30.11.2010	78-37-5	201-110-3	linalyl cinnamate						
2		30.11.2010	90-50-6	201-999-8	3,4,5-trimethoxycinnamic acid						

Default
 Identifiers
 Datasets
 Models
 Endpoints
 All descriptors
 pKa
 Molecule size
 Electronic descriptors (PM3 optimized structure)
 Electronic descriptors (original structure)
 Toxtree: Cramer rules
 http://apps.idea

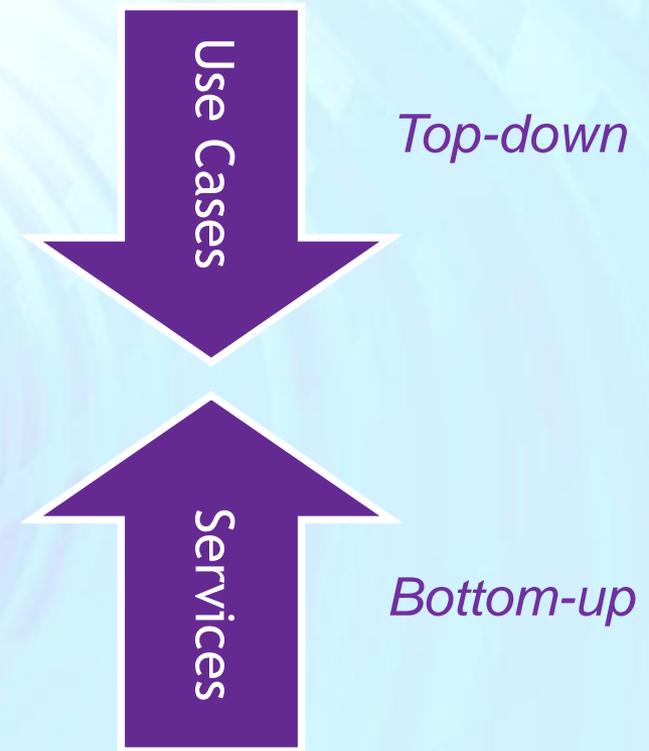
OpenTox: Databases

Dataset	OK	Probably OK	Probably ERROR	Unknown	Probably ERROR%
ECHA list of pre-registered substances	N/A	N/A	N/A	N/A	N/A
Chemical Identifier Resolver	67779	5314	3638	3471	4.75%
ChemIDplus	64802	7986	921	1745	1.24%
ChemDraw	17918	1147	502	478	2.57%
JRC PRS list	61332	4833	4022	2880	5.83%
ISSCAN	931	50	98	62	9.40%
CPDBAS	778	37	0	693	0%
DBPCAN	60	2	0	147	0%
EPAFHM	281	5	0	331	0%
KIERBL	102	1	0	175	0%
IRISTR	346	16	0	177	0%
FDAMDD	213	19	1	983	0.08%
ECETOC skin irritation	158	12	0	5	0%
Skin sensitisation (LLNA)	160	7	4	38	1.95%
Bioconcentration factor (BCF) Gold Standard Database	N/A	N/A	N/A	N/A	N/A

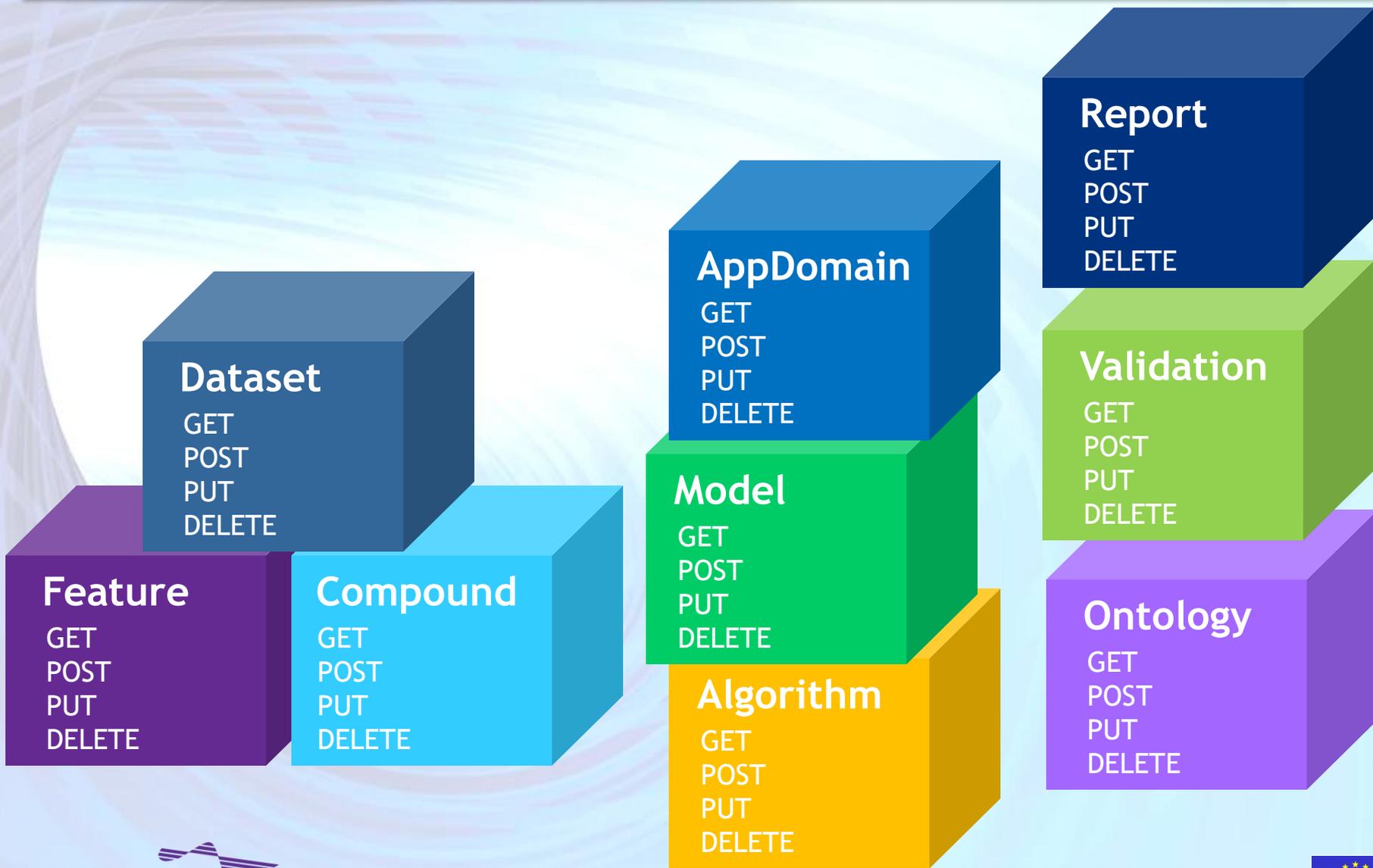
Integrating diverse data sources into OpenTox ontology

Development Process

- Bidirectional
 - from the services (building blocks) to the use cases, and, vice versa,
 - from the use cases to the services
- Use cases
 - ToxPredict (shown later)
 - ToxCreate (shown later)
 - Read-across
 - ToxCast
 - ...



Overview of Application Programming Interfaces



Interface Definitions

Description	Method	URI	Parameter	Result	Status codes
Get available feature URIs for a compound	GET	/compound/{cid}/feature	?feature_uris[]="URI to features" (optional)	Returns representation of the features as uri-list or RDF All available features are returned, if no parameter is specified.	200,404,503
Create a new feature value	POST	/compound/{cid}/feature	?feature_uri="URI to feature" (mandatory, single feature)&value=the_value	URI of the compound with the new feature, e.g. /compound/{id}?feature_uris[]=the-new-feature	200,400,503
Update a new feature value	PUT	/compound/{cid}/feature	?feature_uri="URI to feature" (mandatory, single feature)&value=the_value		200,400,404,503
Delete specified features from the compound	DELETE	/compound/{cid}/feature	?feature_uris[]="URI to features" (optional)		200,400,404,503

Interface Definitions

Description	Method	URI	Parameters	Result	Status codes
get description of a specific feature definition	GET	/feature/{id}	-	URI-list or RDF representation of a feature.	200,404,503
create a new feature	POST	/feature	Content-type ="any-of-RDF-types", content=RDF-representation	URI of the new feature definition.	200,400,404,503
update feature	PUT	/feature/{id}	Content-type ="any-of-RDF-types", content=RDF-representation	-	200,400,404,503
delete feature	DELETE	/feature/{id}	-	-	200,400,404,503
get a list of available feature definitions	GET	/feature	? query =URI-of-the-owl:sameAs-entry	URI list or RDF of features found by the query or all available, if query is empty. Returns all features, for which owl:sameAs is given by the query.	200,404,503

Interface Definitions

Description	Method	URI	Parameters	Result	Status codes
Get a list of available datasets	GET	/dataset	Query parameters (optional, to be defined by service providers).	List of URIs or RDF for the metadata only.	200,404,503
Get a dataset	GET	/dataset/{id}	-	Representation of the dataset in a supported MIME type.	200,404,503
Query a dataset	GET	/dataset/{id}	compound_uris[] and/or feature_uris[] to select compounds and features; further query parameters may be defined by service providers.	Representation of the query result in a supported MIME type.	200,404,503
Get metadata for a dataset	GET	/dataset/{id}/metadata	-	Representation of the dataset metadata in a supported MIME type.	200,404,503
Get a list of all compounds in a dataset	GET	/dataset/{id}/compounds	-	List of compound URIs.	200,404,503
Get a list of all features in a dataset	GET	/dataset/{id}/features	-	RDF or List of feature URIs (pointing to feature definitions/ontologies).	200,404,503

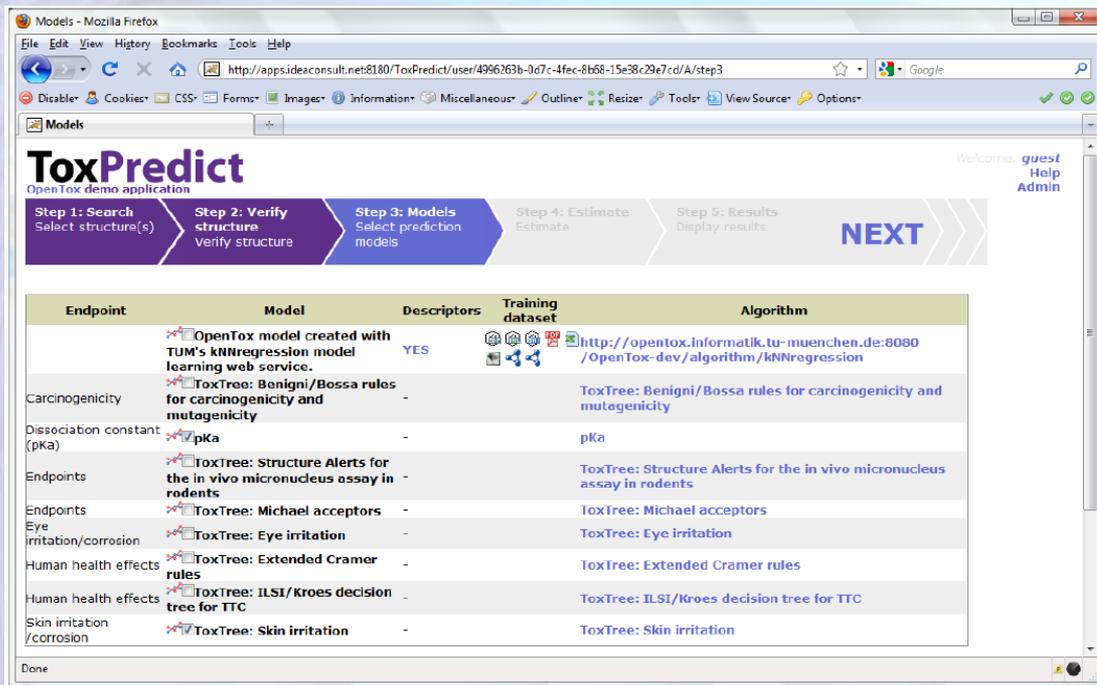
Interface Definitions

Description	Method	URI	Parameters	Result	Status codes
Get a list of all available models	GET	/model	(optional) ?query=URI-of-the-owl:sameAs-entry	List of model URIs or RDF representation. If query specified, returns all models, for which owl:sameAs is given by the query.	200,404,503
Get the representation of a model	GET	/model/{id}	-	Representation of the model in a supported MIME type.	200,404,503
Delete a model	DELETE	/model/{id}	-	-	200,404,503
Apply a model to predict a dataset	POST	/model/{id}	dataset_uri result_dataset=dataseturi dataset_service=datasetseviceuri	URI of created prediction dataset (predictions are features), task URI for time consuming computations.	200,202,400,404,500,503
Apply a model to predict a compound	POST	/model/{id}	compound_uri	Prediction in a supported MIME type; task URI for time consuming computations.	200,202,400,404,500,503

Interface Definitions

Description	Method	URI	Parameters	Result	Status codes
Retrieve SPARQL query results	GET	/ontology	? query =SPARQL_QUERY (mandatory)	RDF representation of the query results.	200,404,500
Predefined query to retrieve all models	GET	/ontology/models		RDF representation of all models.	
Predefined query to retrieve all endpoints	GET	/ontology/endpoints		RDF representation of all endpoints.	
Predefined query to retrieve all algorithms	GET	/ontology/algorithms		RDF representation of all algorithms.	
Submit SPARQL query and/or OpenTox service URL	POST	/ontology	uri []=URL of a OpenTox RDF resource query =SPARQL_QUERY	RDF representation of the query results, if query is specified. if uri [] is specified, the server retrieves a RDF representation and adds it to the RDF storage, thus making it available for the subsequent queries.	200,404,500,502

What you can do with it ...



ToxPredict
OpenTox demo application

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

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

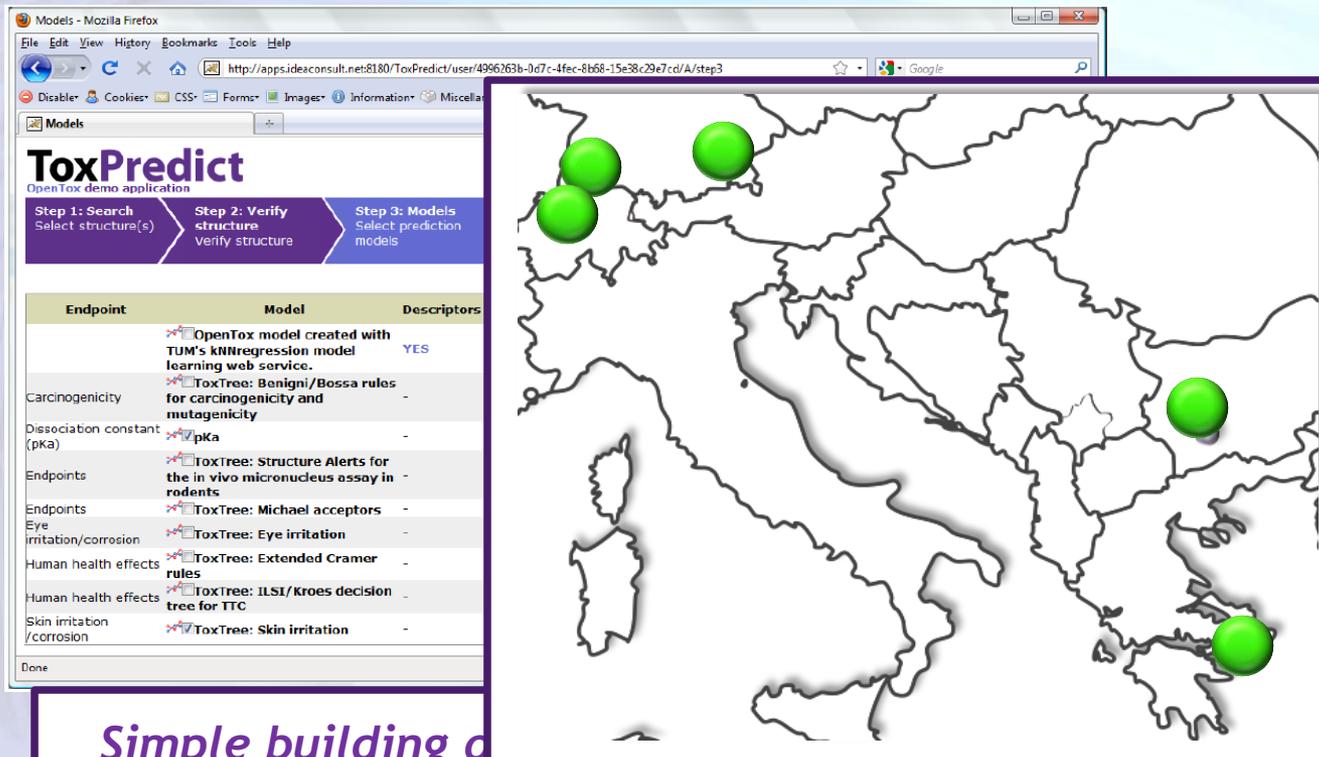
NEXT

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

Done

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 interface in a Mozilla Firefox browser window. The URL is <http://apps.ideaconsult.net:8180/ToxPredict/user/496263b-0d7c-4fec-8b68-15e38c29e7cd/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 the navigation bar is a table with columns for Endpoint, Model, and Descriptors.

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

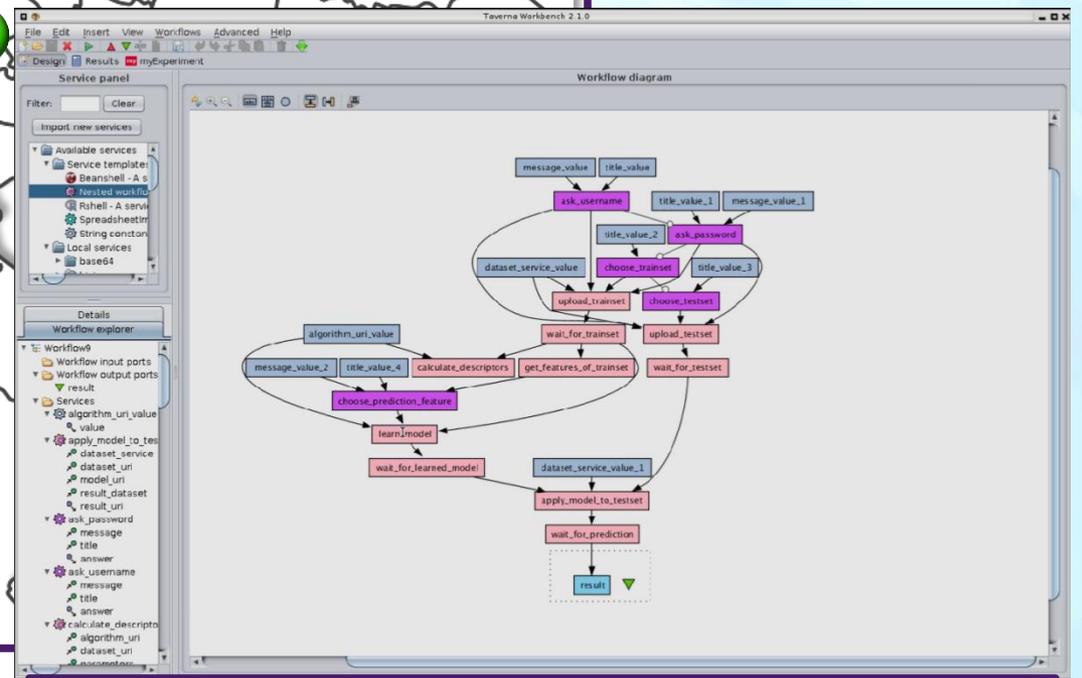
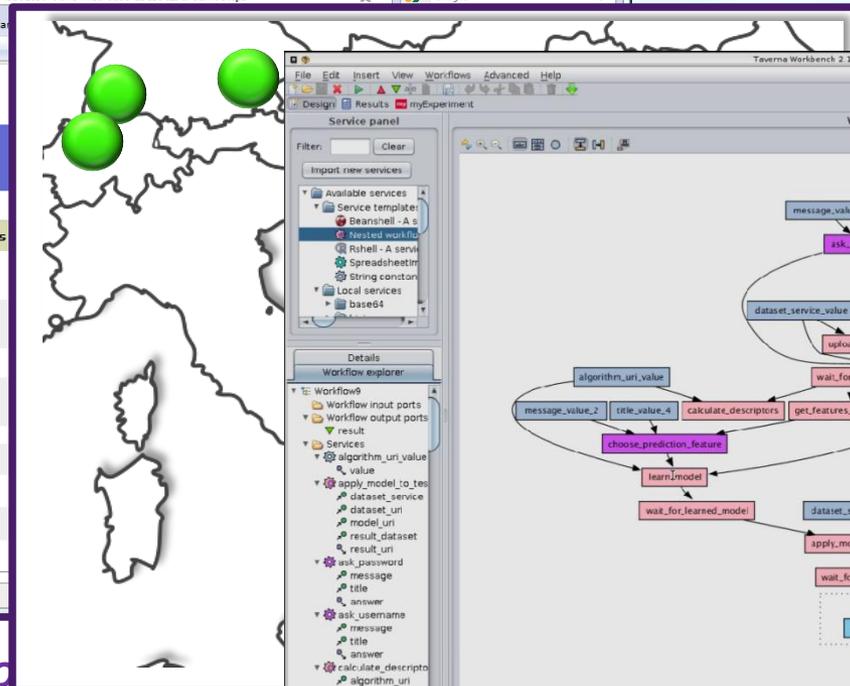
Overlaid on the right side of the screenshot is a map of Europe with five green circular markers placed in various locations across the continent.

Simple building of applications methods and

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

What you can do with it ...

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



Simple building of applications methods and

Distributed of wide range of methods

Integration into workflow systems for computational biology

ToxCast

- Focus on prediction of toxicological endpoints
- Use *in vitro* data to predict *in vivo* endpoints
- Multi label approach
 - approximately 400 labels, 300 instances, 1600 features

Structure	<i>In vitro</i>			<i>In vivo</i>		
	ncgc_ar_agonist	ncgc_ar_antagonist	...	chr_rat_thyroid	ch_rat_liver	...
Abamectin	1	0	...	0	0	...
Acephate	0	0	...	0	1	...
Acetamiprid	1	1	...	1	0	...
...

ToxCast: Approach and Further Steps

- Use all data available at this point, including data points with **missing values**
- Take into account interdependencies between *in vivo* endpoints (**multi-label classification**)
 - improves upon predicting endpoints individually
- Use **applicability domain** on *in vitro* data to avoid unjustified predictions
 - improves upon predicting everything
- Results comparable to other multi-label datasets
- Further steps: take into account descriptions of assays to link into biological databases

Taverna Workflow System

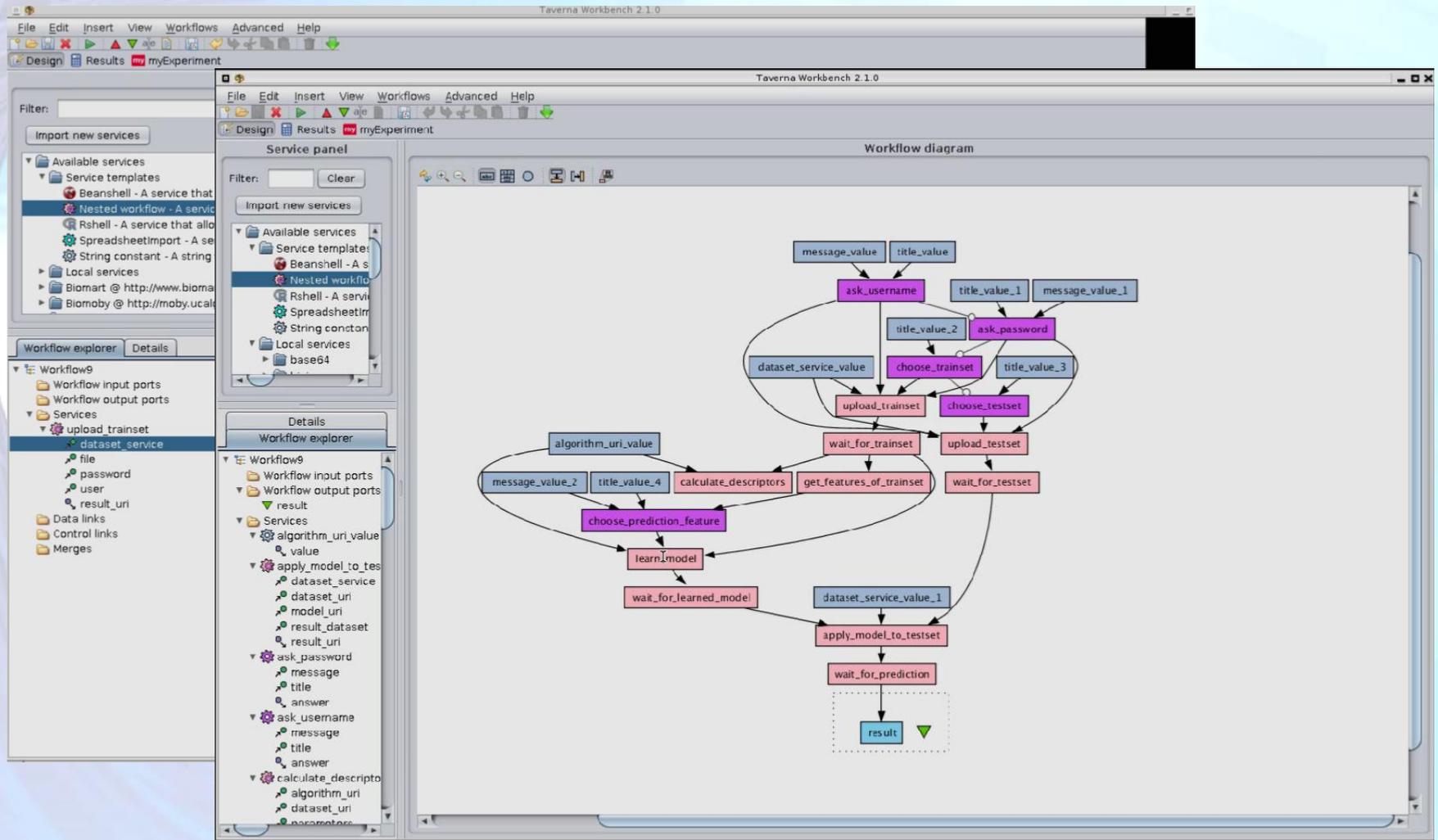
- Standard for workflows in computational biology
- OpenTox services can be integrated into Taverna workflows
- Allows for the integration of biological information
 - pathway data, ...
 - mechanistic explanation

Taverna Workflow System

The screenshot displays the Taverna Workbench 2.1.0 interface. The main window is divided into several panels:

- Service panel:** Contains a list of available services under 'Service templates' and 'Local services'. The 'Nested workflow' service is currently selected.
- Workflow diagram:** Shows a workflow graph with nodes: 'dataset_service', 'password', 'user', 'file', 'query_server', and 'result_uri'. Arrows indicate the flow of data between these components.
- Input dialog:** A small window titled 'Input' is open, prompting the user to 'Enter string value'. The text '8080/ambit2/dataset/291/features' is entered in the input field.
- Workflow explorer:** Located at the bottom left, it shows a tree view of the workflow structure, including 'Workflow9', 'Workflow input ports', 'Workflow output ports', 'Services', and 'upload_trainset'.

Taverna Workflow System



Consequences for Requirements on OpenTox

User Requirements		Software Requirements
Unambiguous data	⇒	<i>formal way of representing information about data</i>
Unambiguous access	⇒	<i>well-defined interfaces</i>
Transparency of computational tools	⇒	<i>formal way of representing information about methods, well-defined interfaces</i>
Variety of user groups	⇒	<i>simplicity and modularity of design</i>
Need to integrate various resources (e.g., databases, prediction methods, models, ...) to make meaningful predictions	⇒	<i>distributed architecture, interoperability</i>
Need to integrate biological information	⇒	<i>again, modularity of design, extensibility</i>

Strategic Context/Goals

- REACH: possible reduction of test animals by using existing experimental data in conjunction with QSAR
- Also practical needs: **reporting** and form filling
- By the OECD principles, a number of requirements to a framework like OpenTox arise

	OECD Principle	OpenTox addresses 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 prediction of toxicological mechanisms and the recording of opinions and analysis in reports

Summary and Future Work

- Comprehensive framework for predictive toxicology that allows you to address a wide range of tasks
 - interface definitions, services (also for: validation, **reporting**, ...), use cases
 - **interoperability** and **extensibility by design**
- Further work on other use cases: e.g.
 - Read Across (technical infrastructure available)
 - ToxCast
 - pathway prediction service: developing a service for predicting transformation products
 - Synergy pilot
 - ...

OpenTox Application Demonstrations

Nina Jeliazkova (Ideaconsult, Bulgaria)

Christoph Helma (In Silico Toxicology, Switzerland)

Andreas Karwath (Albert-Ludwigs Univ. Freiburg, Germany)

Use Cases

ToxPredict

- offers easy access for toxicological hazard estimation of a chemical structure
- for non-QSAR specialists
- a simple yet easy-to-use user interface

ToxCreate

- is aimed at researchers in
- life sciences and toxicology
 - QSAR experts
 - people interested in machine learning/statistics
 - pharmaceutical industry R&D

ToxPredict

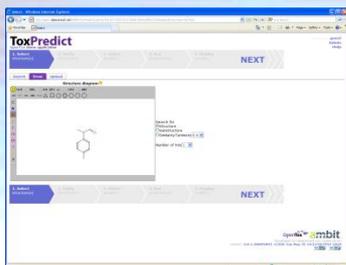
The use case can be divided into the following five steps:

1. Enter/select a chemical compound
2. Display selected/found structures
3. Select models
4. Perform the estimation
5. Display the results

→ Live Demo: www.toxpredict.org

ToxPredict: Step 1 (behind the scenes)

Find structure by name, registry number, SMILES, InChI, structure, substructure, similarity...



ToxPredict
Web
Application

OT Dataset API *HTTP GET*

OT Ontology
Service

text/uri-list,
application/rdf+xml,
chemical/x-daylight-smiles
chemical/x-mdl-sdfile,...

Here is the list of structures as
URI links, RDF, MOL or SMILES.



ToxPredict: Step 3 (behind the scenes)



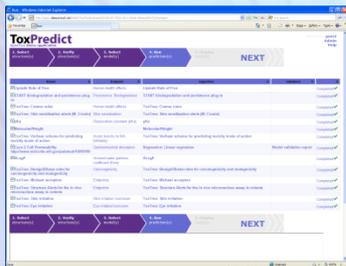
ToxPredict
Web
Application

What prediction models are available? Is there a model for endpoint X?

HTTP GET SPARQL query

OT Ontology
Service

application/sparql-results+xml



Model URI	Related Endpoints	Algorithms
OT:0000001	OT:0000001	OT:0000001
OT:0000002	OT:0000002	OT:0000002
OT:0000003	OT:0000003	OT:0000003
OT:0000004	OT:0000004	OT:0000004
OT:0000005	OT:0000005	OT:0000005
OT:0000006	OT:0000006	OT:0000006
OT:0000007	OT:0000007	OT:0000007
OT:0000008	OT:0000008	OT:0000008
OT:0000009	OT:0000009	OT:0000009
OT:0000010	OT:0000010	OT:0000010

Here is the list of model URIs and related endpoints and algorithms in SPARQL format.

ToxCreate: intended audience

Toxicologists with

- access to toxicological data
- basic computer skills
- little or no knowledge of QSAR algorithms

Goal:

- build and validate prediction models from user provided training data

→ Live demo www.toxcreate.org

Future Development

Inclusion of further QSAR algorithms

- build and validate multiple models
- automated selection of the best model (based on validation results)

GUI refinement:

- based on user feedback

Expert interface:

- access to all model building parameters
- command line/GUI versions

Behind the Scenes of ToxCreat

ToxCreat

Create Inspect Predict About

This service is for testing purposes only - once a week all models will be deleted. Please send bug reports and feature requests to our **issue tracker**.

This service creates **lazar** classification models (more model building algorithms will follow) from your uploaded datasets. Here are **instructions** , for creating training datasets in Excel.

ToxCreat

Create Inspect Predict About

This service is for testing purposes only - once a week all models will be deleted. Please send bug reports and feature requests to our **issue tracker**.

This service creates **lazar** classification models (more model building algorithms will follow) from your uploaded datasets. Here are **instructions** , for creating training datasets in Excel.

1. Enter a name for your endpoint:

2. Upload training data in **CSV** format:

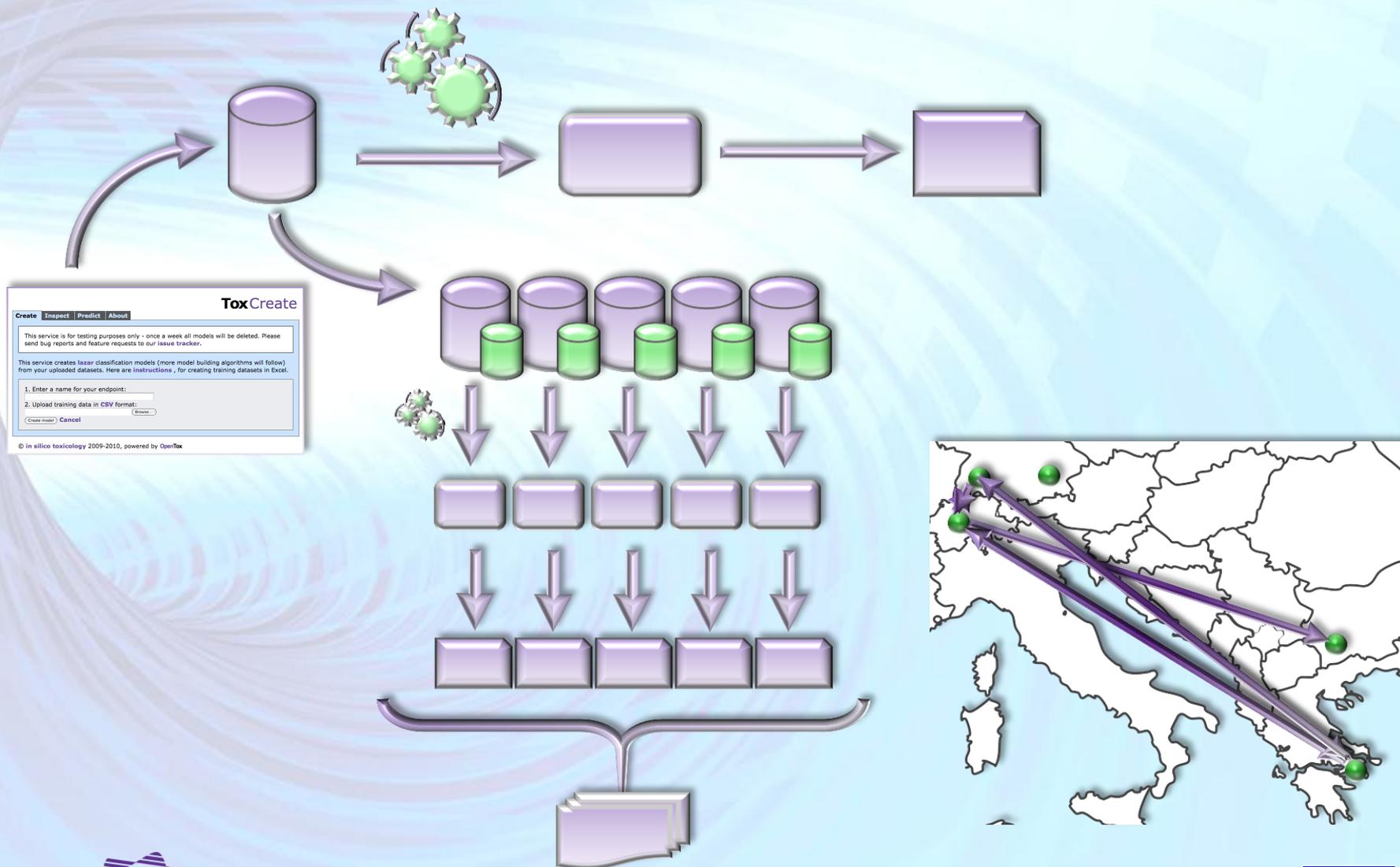
© in silico toxicology 2009-2010, powered by OpenTox

1. Enter a name for your endpoint:

2. Upload training data in **CSV** format:

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

Behind the Scenes of ToxCreat



Current State of the System

Web services online:

AlgorithmWS: NTUA, TUM, IDEA, IST

ModelWS: NTUA, TUM, IDEA, IST

FeatureWS: NTUA, TUM, IDEA, IST, ALU-FR

CompoundWS: NTUA, TUM, IDEA, IST

ValidationWS: ALU-FR

DatasetWS: NTUA, TUM, IDEA, IST, ALU-FR

Collaboration, Sustainability & Future Directions

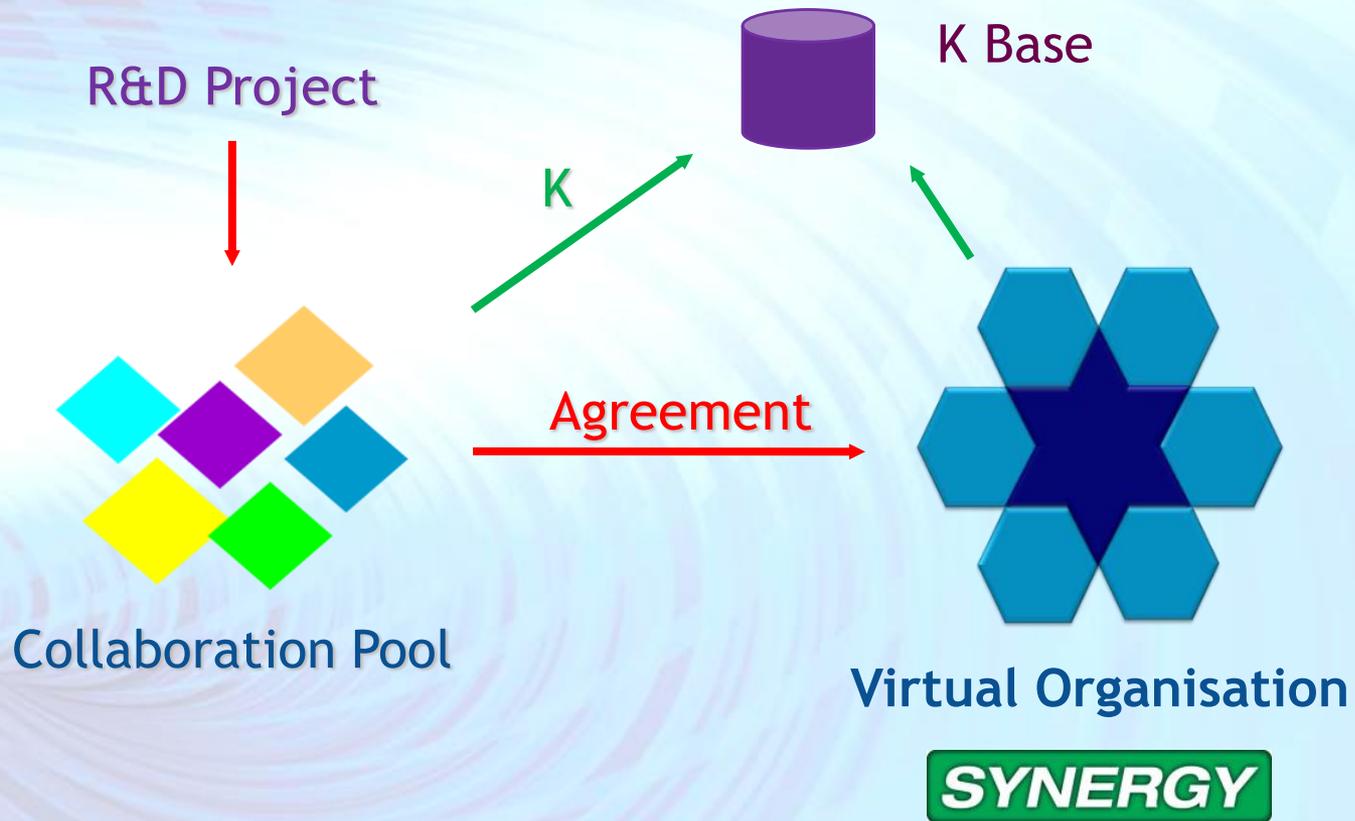
Barry Hardy
Douglas Connect
OpenTox Project Coordinator

Our Drivers - Taking on Technical, Cultural and “Other” Challenges of the Unexpected

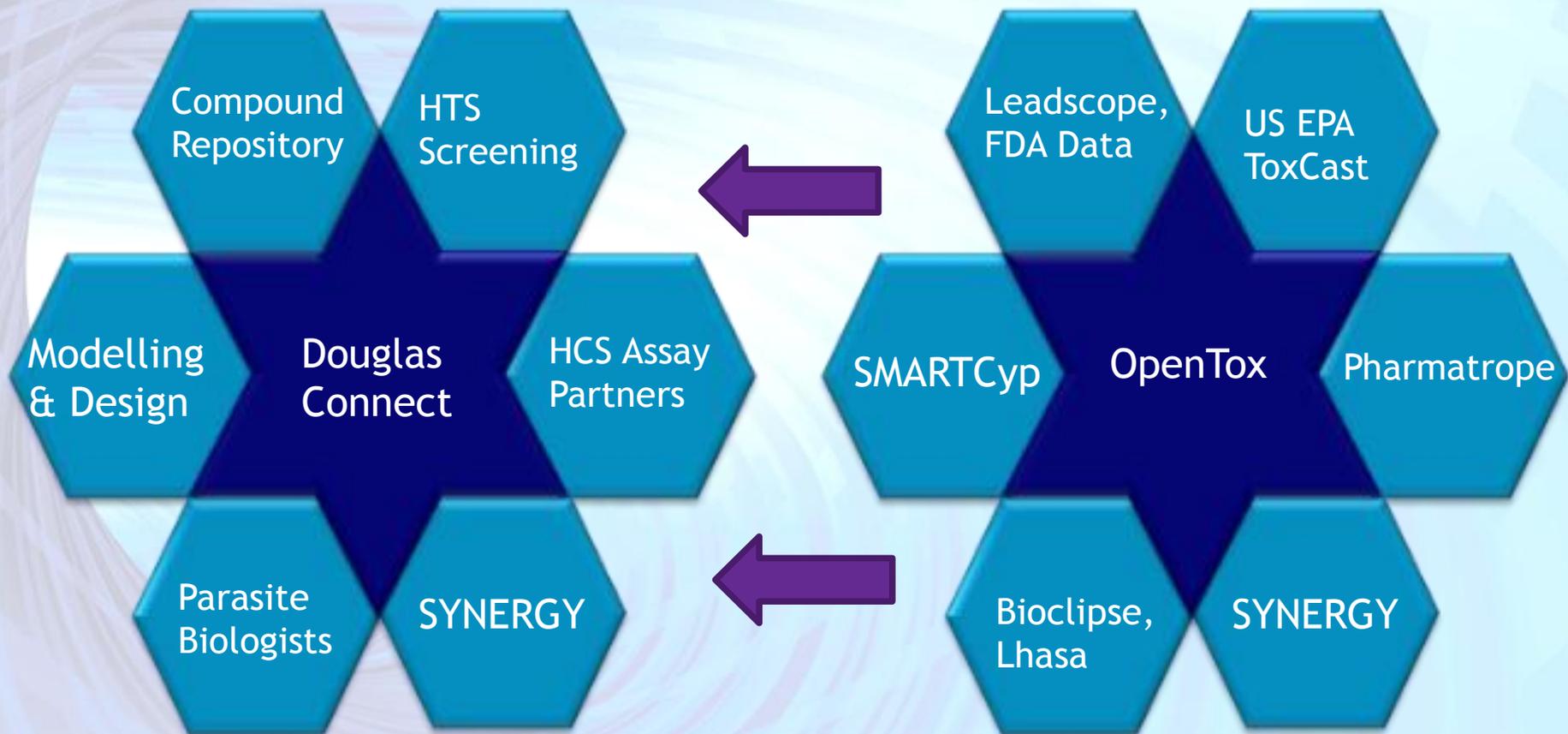


[Visit with Lions at Mukuni Reintroduction Project, Livingstone, Zambia](#)

Virtual Organisation Pilots



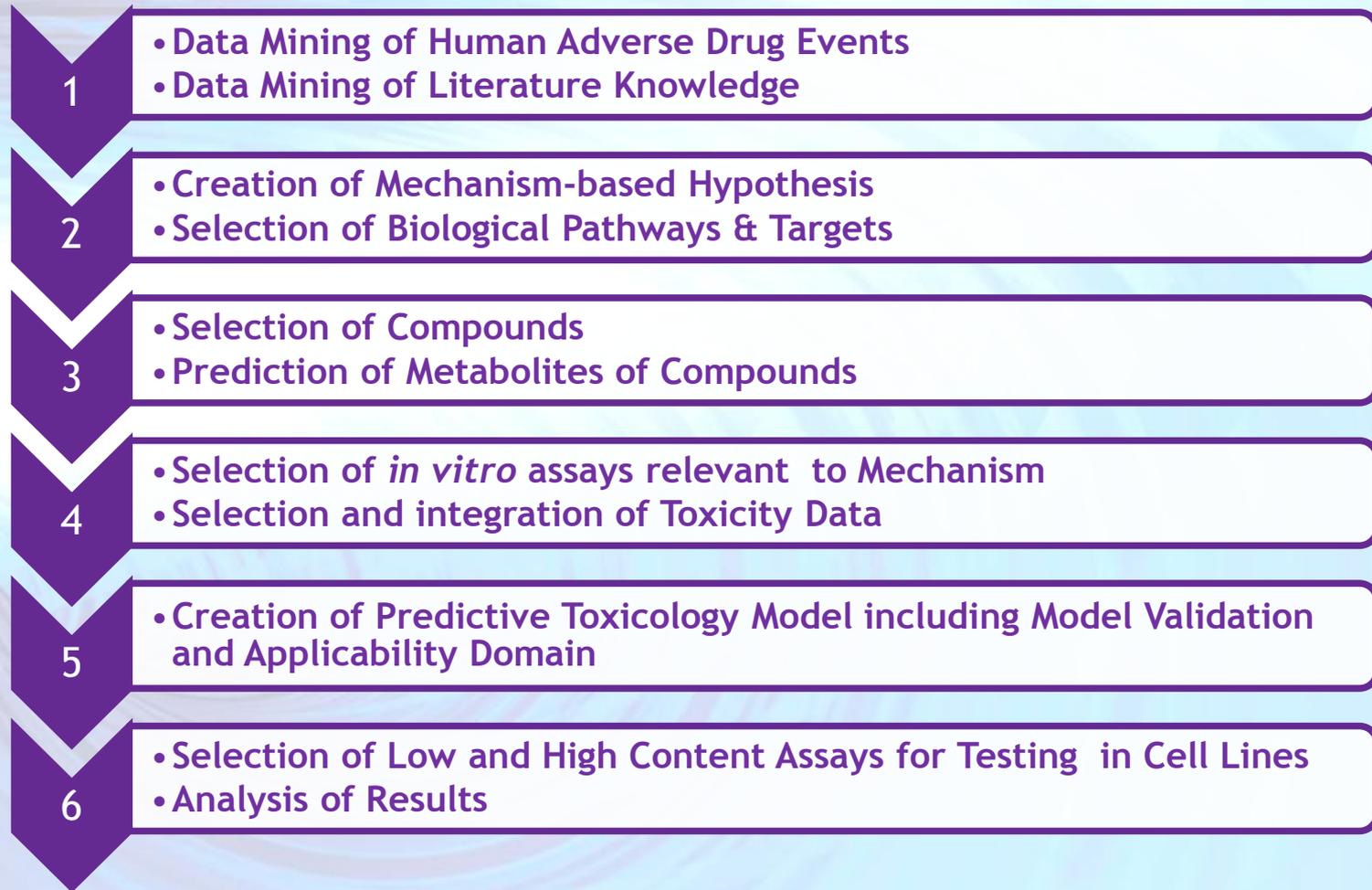
Virtual Organisation Pilots



Neglected Disease Drug Design VO

Predictive Toxicology VO

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



Recording of Collaborative R&D

Controlled Vocabularies

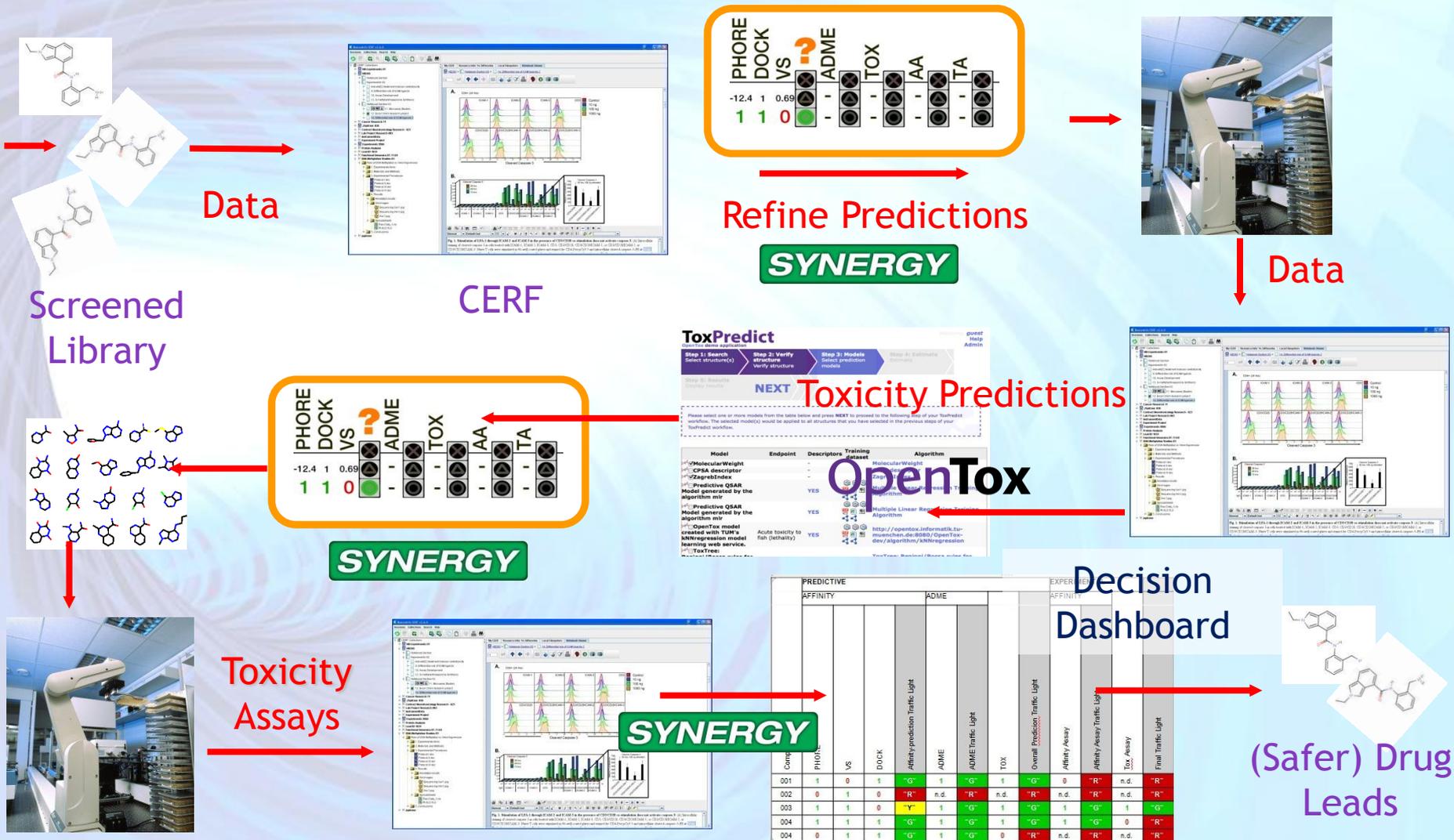
Visualisation

Collaborative Electronic Laboratory Notebook (ELN)

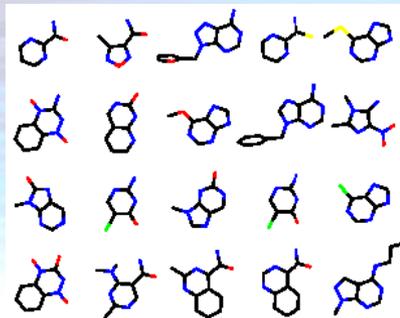
Fig. 1. Stimulation of LFA-1 through ICAM-2 and ICAM-3 in the presence of CD3/CD28 co-stimulation does not activate caspase-3. (A) Intracellular staining of cleaved-caspase 3 in cells treated with ICAM-1, ICAM-2, ICAM-3, CD3, CD3/CD28, CD3/CD28/ICAM-1, or CD3/CD28/ICAM-2, or CD3/CD28/ICAM-3. Naive T cells were stimulated in 96-well coated plates and stained for CD4-PerpCy5.5 and intracellular cleaved-caspase-3-FE at 24 hrs.

Condition	24 hrs	48 hrs	72 hrs
IgG	~100	~100	~100
ICAM-1	~100	~100	~100
ICAM-2	~100	~100	~100
ICAM-3	~100	~100	~100
CD3	~100	~100	~100
CD3/CD28	~100	~100	~100
CD3/CD28/ICAM-1	~100	~100	~100
CD3/CD28/ICAM-2	~100	~100	~100
CD3/CD28/ICAM-3	~100	~100	~100

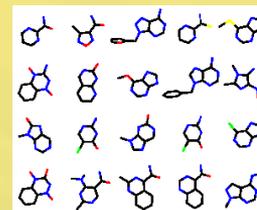
Synergy Collaboration Pilots



1. A library of compounds is entered to the ELN



ELN

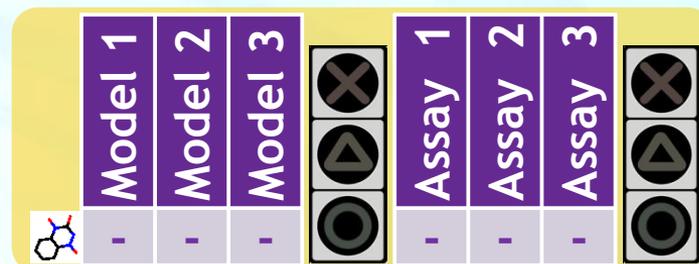


Synergy

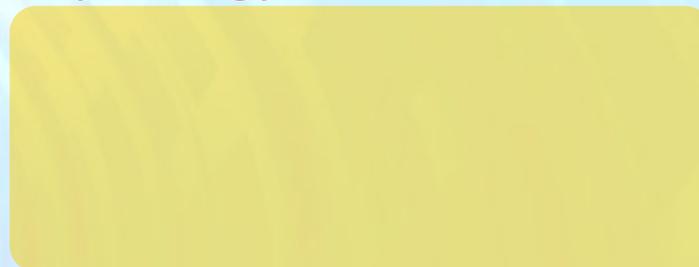
OpenTox

2. Each compound is assigned a data structure in ELN

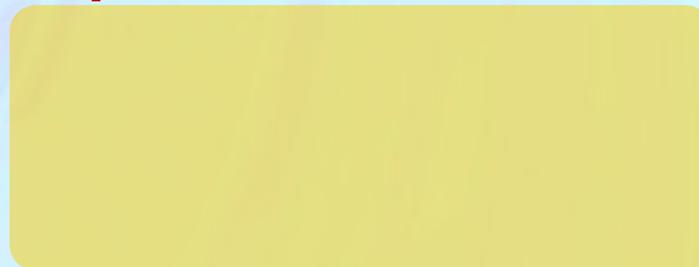
ELN



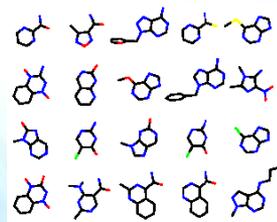
Synergy



OpenTox



3. ELN passes compounds to OpenTox and SYNERGY



ELN

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

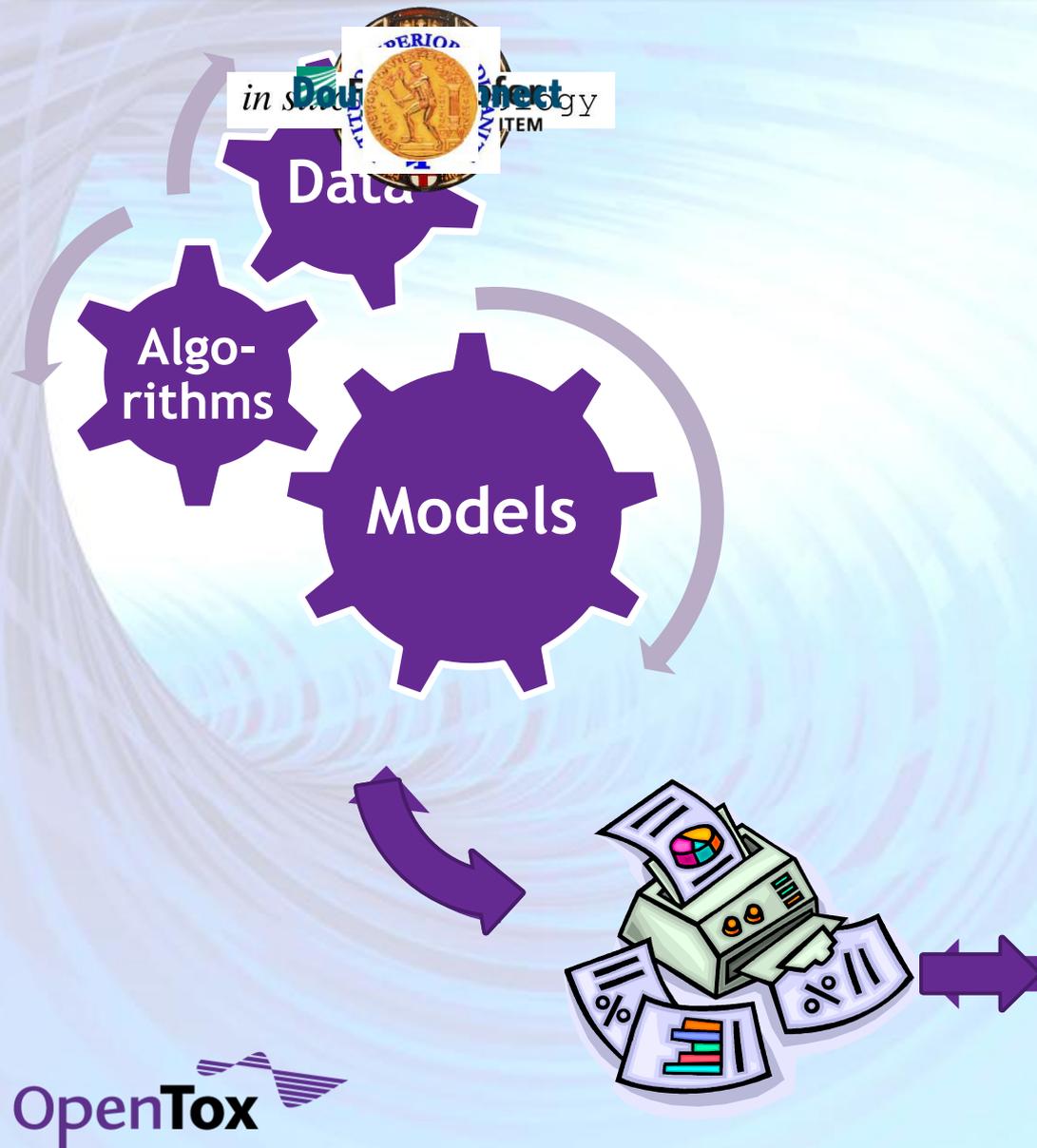
Synergy

	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		
-	-	-	-	-	-	-	-	-	-	-

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	<input type="checkbox"/>	Assay 1	Assay 2	Assay 3	<input type="checkbox"/>
	1	0	1	<input type="checkbox"/>	-	-	-	<input type="checkbox"/>

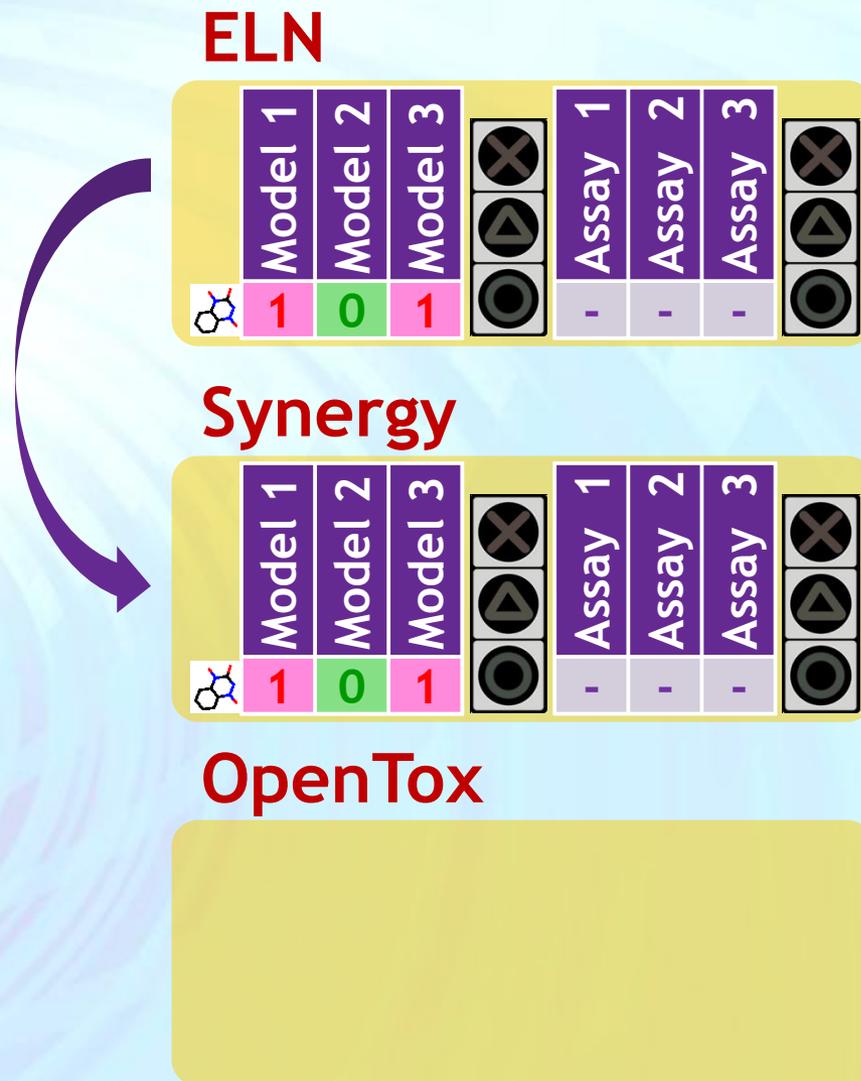
Synergy

	Model 1	Model 2	Model 3	<input type="checkbox"/>	Assay 1	Assay 2	Assay 3	<input type="checkbox"/>
	-	-	-	<input type="checkbox"/>	-	-	-	<input type="checkbox"/>

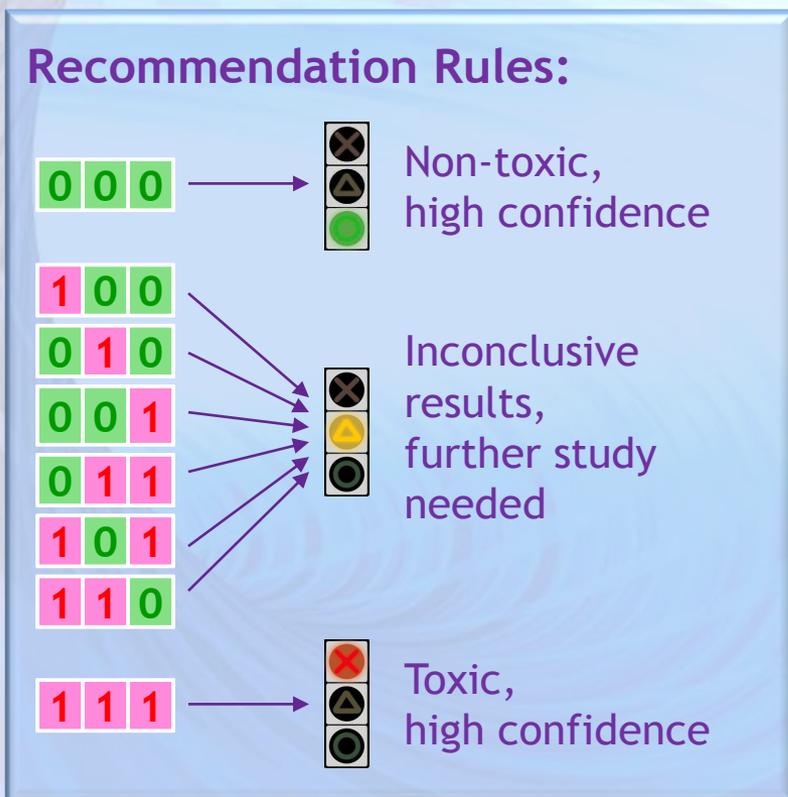
OpenTox

	Model 1	Model 2	Model 3	<input type="checkbox"/>	Assay 1	Assay 2	Assay 3	<input type="checkbox"/>
	1	0	1	<input type="checkbox"/>	-	-	-	<input type="checkbox"/>

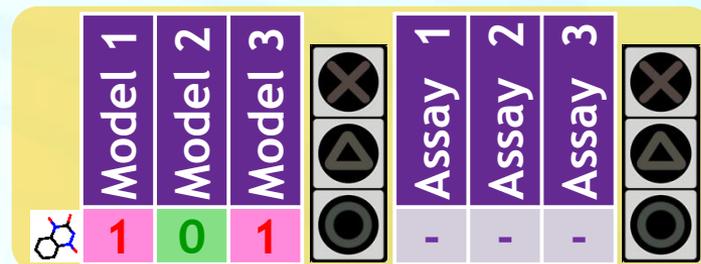
6. ELN sends the results to SYNERGY



7. SYNERGY applies the Recommendation Rules



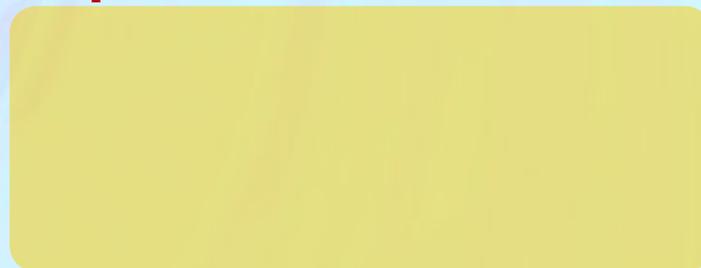
ELN



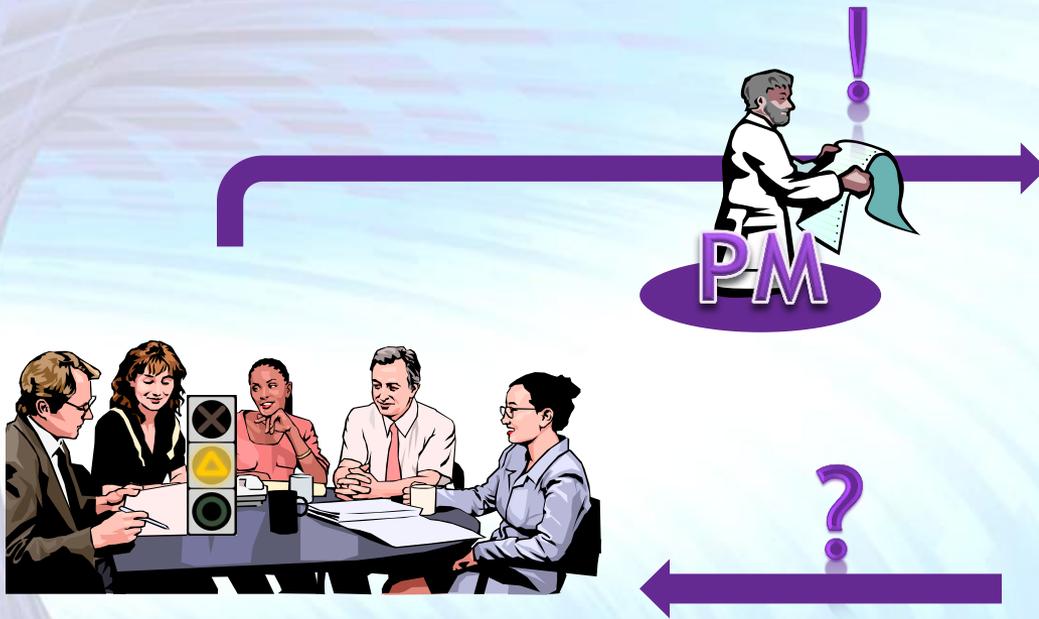
Synergy



OpenTox



8. Inconclusive data → SYNERGY calls a meeting



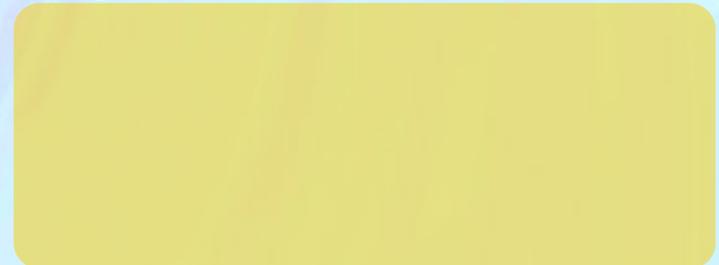
ELN

	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	
	1	0	1		-	-	-	

OpenTox



9. Experimental assays confirm toxicity



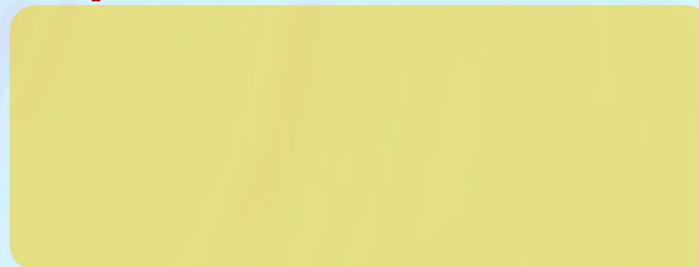
ELN

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

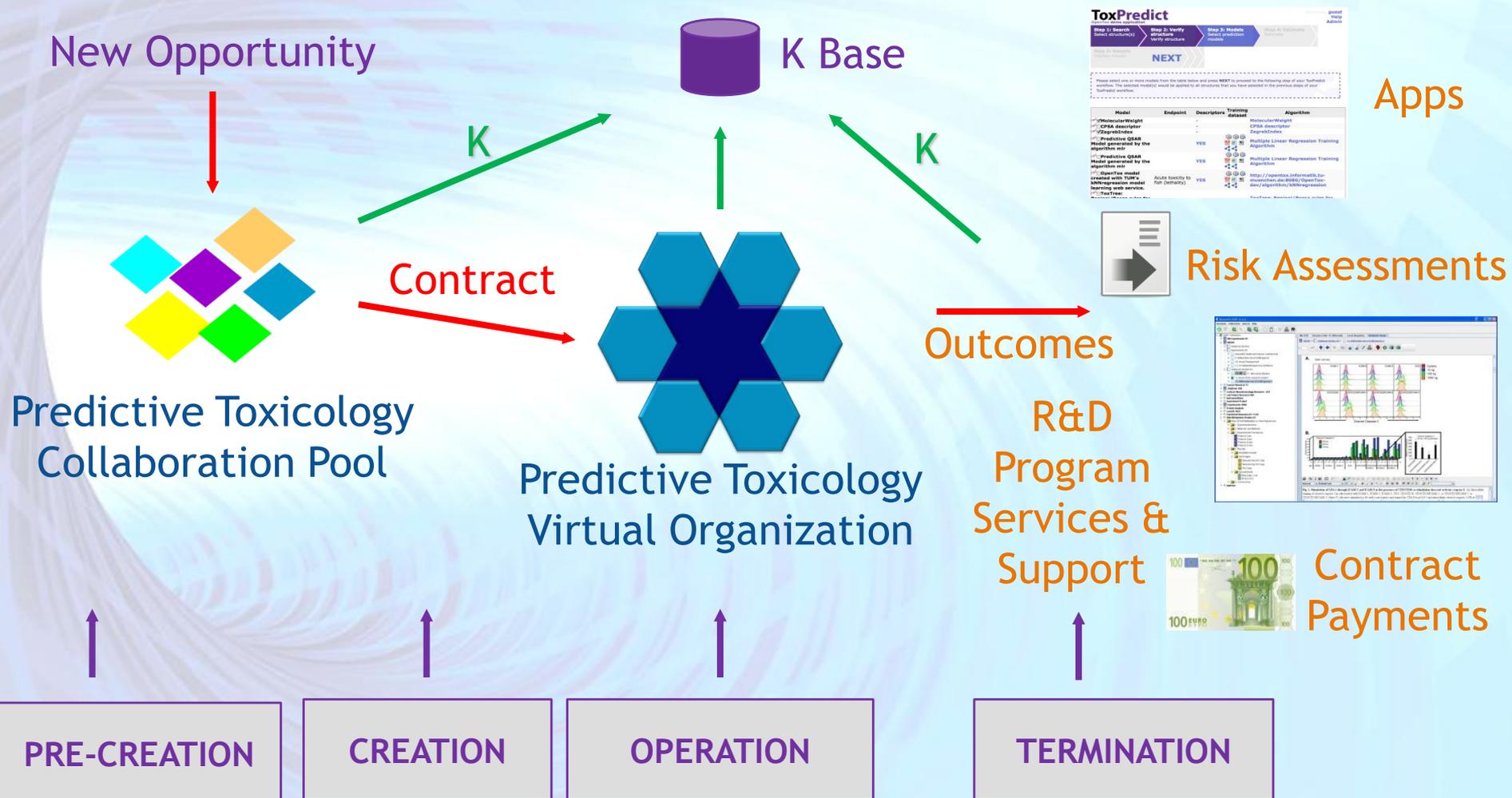
Synergy

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

OpenTox



Sustainability Model



ToxPredict

Step 1: Search
Select structure(s)
Verify structure(s)

Step 2: Verify
Verify structure(s)
Verify prediction
Review

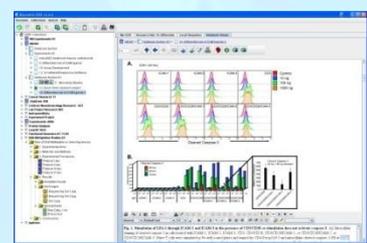
Step 3: Model
Model generation
Model

Step 4: Results
Results

NEXT

Please scroll up to view results from the web browser and click 'NEXT' to proceed to the following step of your 'ToxPredict' workflow. The selected model(s) would be added to the structure(s) that you have selected in the previous steps of your 'ToxPredict' workflow.

Model	Endpoint	Descriptors	Training dataset	Algorithm
Model generated by the algorithm mir	Acute toxicity (fish, 96h)	CPSA descriptor ClogP descriptor	WaterlooMTP100 CPSA descriptor Organics	Multiple Linear Regression Training Algorithm
Model generated by the algorithm mir	Acute toxicity (fish, 96h)	CPSA descriptor ClogP descriptor	WaterlooMTP100 CPSA descriptor Organics	Multiple Linear Regression Training Algorithm
Model generated by the algorithm mir	Acute toxicity (fish, 96h)	CPSA descriptor ClogP descriptor	WaterlooMTP100 CPSA descriptor Organics	Multiple Linear Regression Training Algorithm



Acknowledgements - OpenTox Partners

In Silico Toxicology,
Switzerland

Douglas Connect,
Switzerland

Albert Ludwigs University
Freiburg, Germany

Ideaconsult,
Bulgaria

Istituto Superiore
di Sanità, Italy

National Technical
University of Athens,
Greece

Technical University
of Munich, Germany



Fraunhofer Institute
for Toxicology &
Experimental Medicine,
Germany

David Gallagher, UK

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

Seascape Learning,
India

Acknowledgements - Co-workers

Barry Hardy

Nicki Douglas

Roman Affentranger

Christoph Helma

Michael Rautenberg

Nina Jeliaskova

Vedrin Jeliaskov

Luben Boyanov

Chelsea Jiang

Martin Martinov

Romualdo Benigni

Olga Tcheremenskaia

Stefan Kramer

Tobias Girschick

Fabian Buchwald

Jörg Wicker

Andreas Karwath

Martin Gütlein

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

Sunil Chawla

Steve Bowlus

Indira Ghosh

Surajit Ray

Gaurav Singhai

Om Prakash

Sylvia Escher

Sara Weiss

Helvi Grimm

OpenTox Advisory Board

- European Centre for the Validation of Alternative Methods
- Pharmatropé
- Bioclipse
- U.S. Environmental Protection Agency
- U.S. Food & Drug Administration
- Nestlé
- Roche
- AstraZeneca
- LHASA
- Leadscope
- University of North Carolina
- EC Environment Directorate General
- Organisation for Economic Cooperation & Development
- CADASTER
- Bayer Healthcare

Final words...

For more information, visit

www.opentox.org

Contact me:

barry.hardy@douglasconnect.com

**Many thanks for your
attention!**



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).