A Brief History of OpenTox

25 June 2015

Barry Hardy (Douglas Connect GmbH)

Barry.Hardy@douglasconnect.com

www.douglasconnect.com

Douglas Connect

Working communities

Communities -> Collaboration -> Innovation

OpenTox

Overview

1) OpenTox Framework and Goals
2) Use Cases & Applications
3) Infrastructure Development
4) Integrated Analysis
5) Safety Assessment
6) Non-Profit Association Formation & Activities
7) Next Steps

Islands

Source: Baily Ed, U.S. Fish and Wildlife Service

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

OpenTox and Open Components and Standards

<-New API addition from ToxBank

Investigation (Study, Assay)
GET
POST
PUT
DELETE

Authorisation & Authentication
GET
POST
PUT
DELETE

Report
GET
POST
PUT
DELETE

AppDomain
GET
POST
PUT
DELETE

Dataset
GET
POST
PUT
DELETE

Validation
GET
POST
PUT
DELETE

Model
GET
POST
PUT
DELETE

Feature
GET
POST
PUT
DELETE

Compound
GET
POST
PUT
DELETE

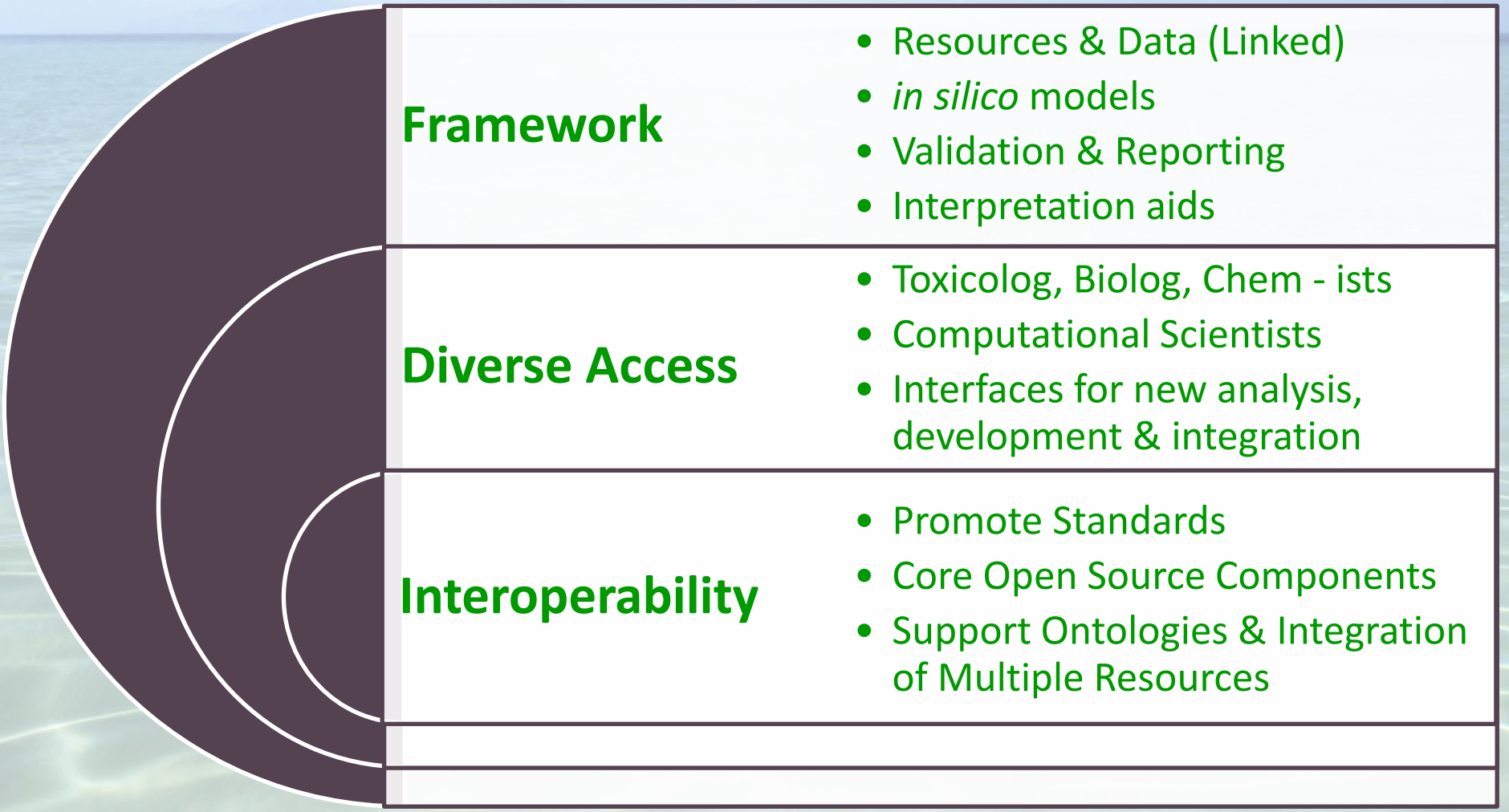
Ontology
GET
POST
PUT
DELETE

Algorithm
GET
POST
PUT
DELETE

OpenTox

www.opentox.org/dev/apis/api-1.2

OpenTox is an Integrating Framework



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

Linked Data is a term used to describe the exposing, sharing, and connecting of data on the Semantic Web using:
URIs a generic means to identify entities in the world
HTTP a simple yet universal mechanism for retrieving resources
RDF a generic graph-based data model with which to structure and link data

Linked Data needs:

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

DBpedia = Linked Data approach applied to Wikipedia

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

The OpenTox Framework (reported 2010)

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

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

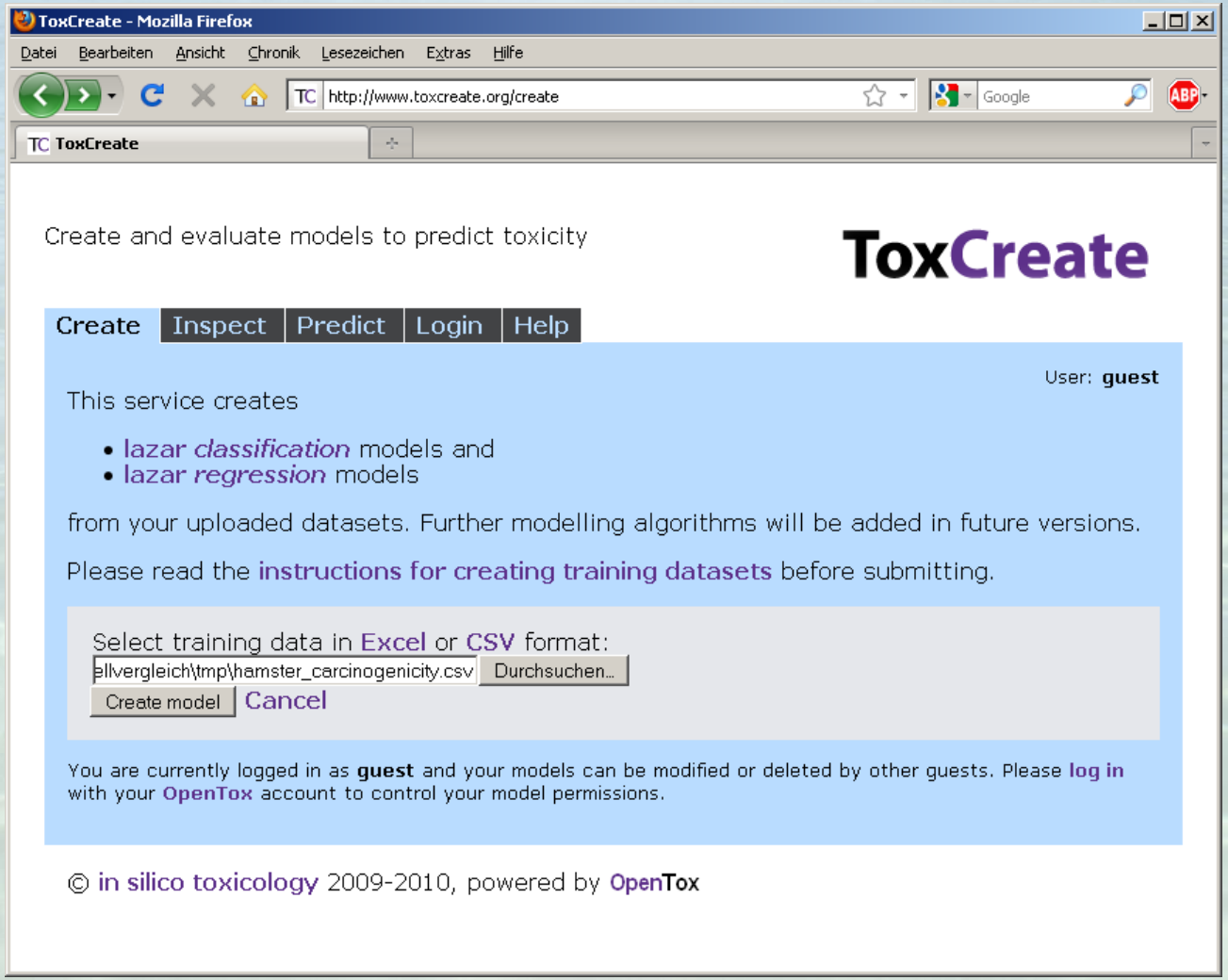
Open Access publication available at
www.jcheminf.com/content/2/1/7

A Toxicology Ontology Roadmap

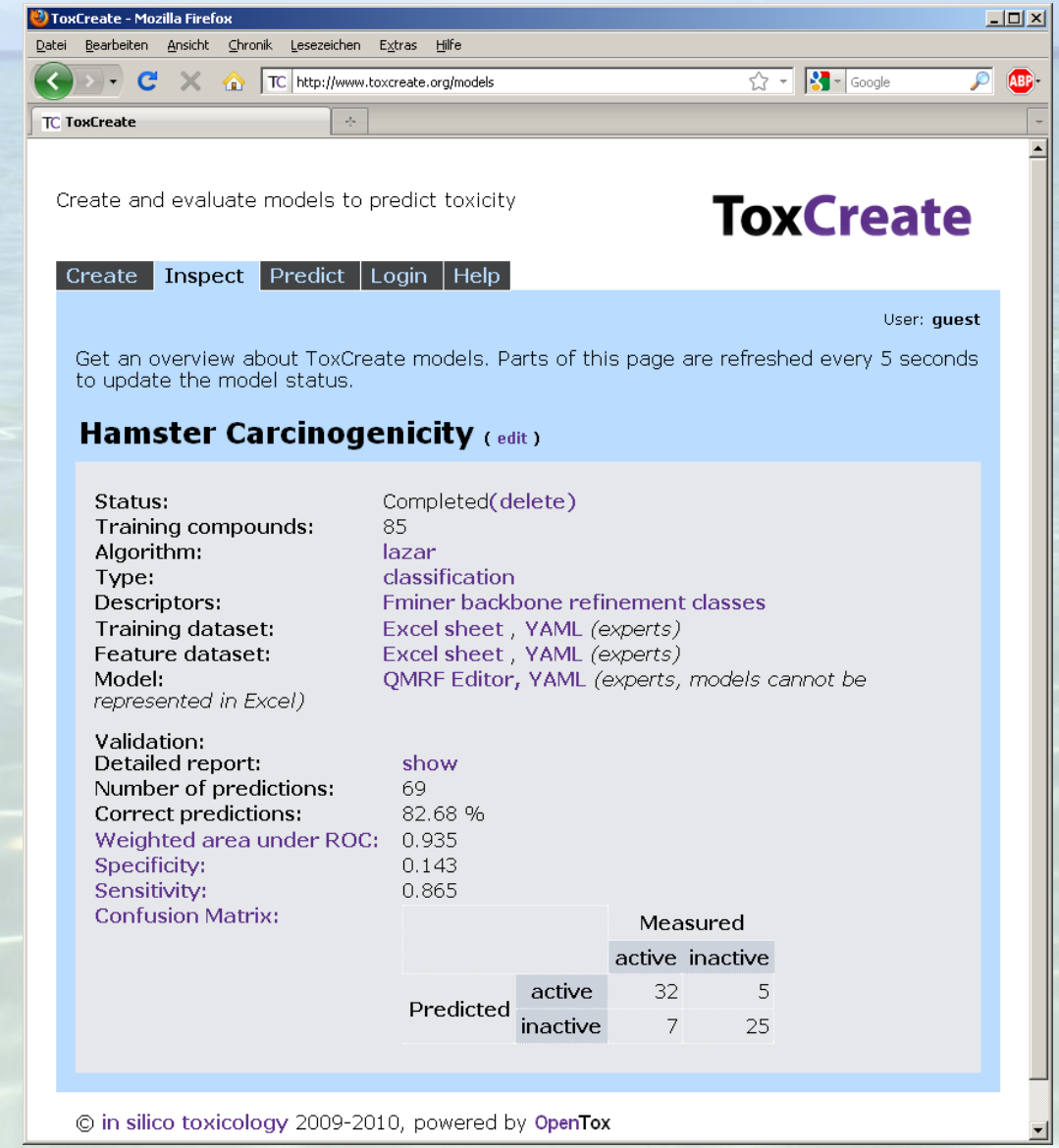
- See perspectives and roadmap published in A Toxicology Ontology Roadmap ALTEX 29(2), 129- 137 and Toxicology Ontology Perspectives 139 - 156 (2012)
- Available online in Open Access mode from www.altex.ch
- Barry Hardy (Douglas Connect and OpenTox), Gordana Apic (Cambridge Cell Networks), Philip Carthew (Unilever), Dominic Clark (EMBL-EBI), David Cook (AstraZeneca), Ian Dix (AstraZeneca & Pistoia Alliance), Sylvia Escher (Fraunhofer Institute for Toxicology & Experimental Medicine), Janna Hastings (EMBL-EBI), David J. Heard (Novartis), Nina Jeliazkova (Ideaconsult), Philip Judson (Lhasa Ltd.), Sherri Matis-Mitchell (AstraZeneca), Dragana Mitic (Cambridge Cell Networks), Glenn Myatt (Leadscope), Imran Shah (US EPA), Ola Spjuth (University of Uppsala), Olga Tcheremenskaia (Istituto Superiore di Sanità), Luca Toldo (Merck KGaA), David Watson (Lhasa Ltd.), Andrew White (Unilever), Chihae Yang (Altamira)

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

ToxCreate - (Q)SAR Model Building application

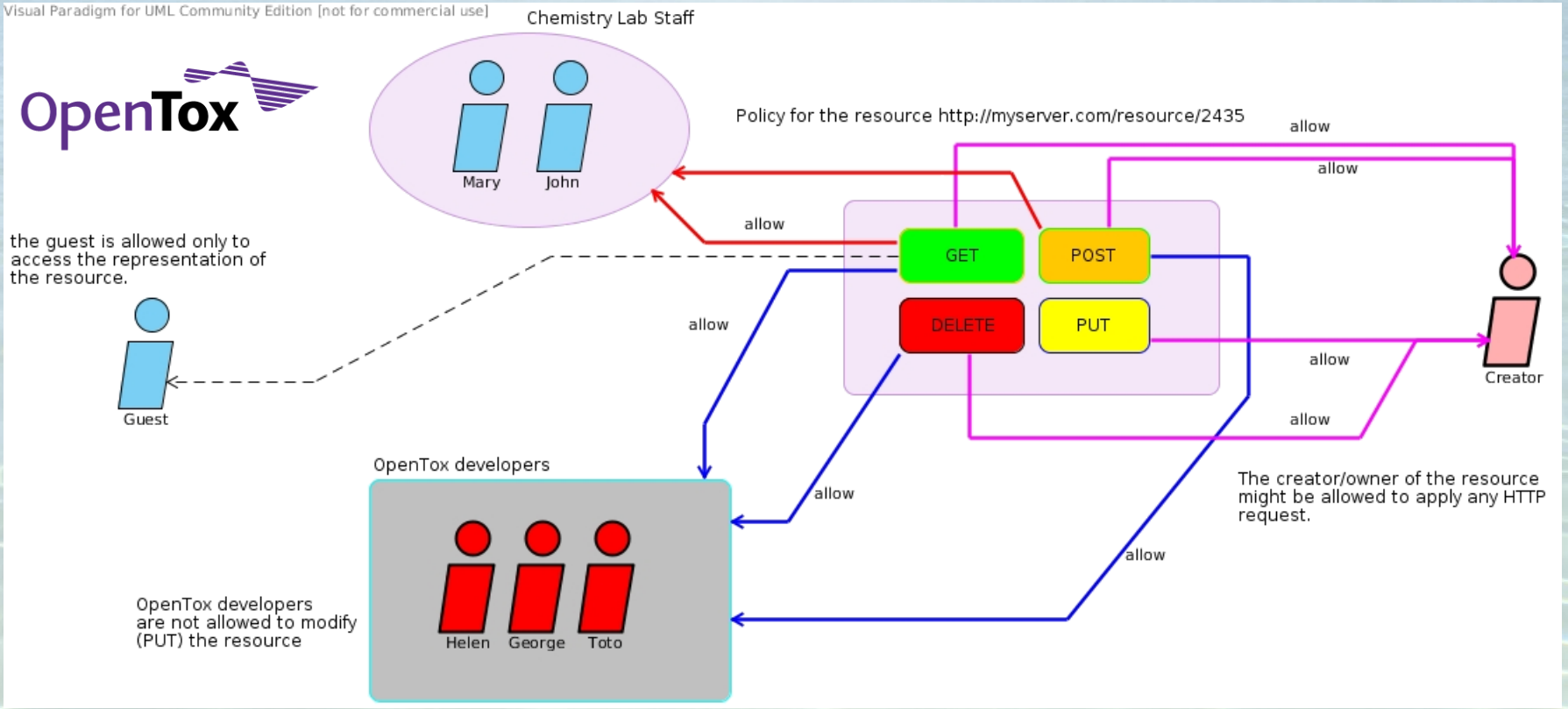


ToxCreate – (Q)SAR Model Results



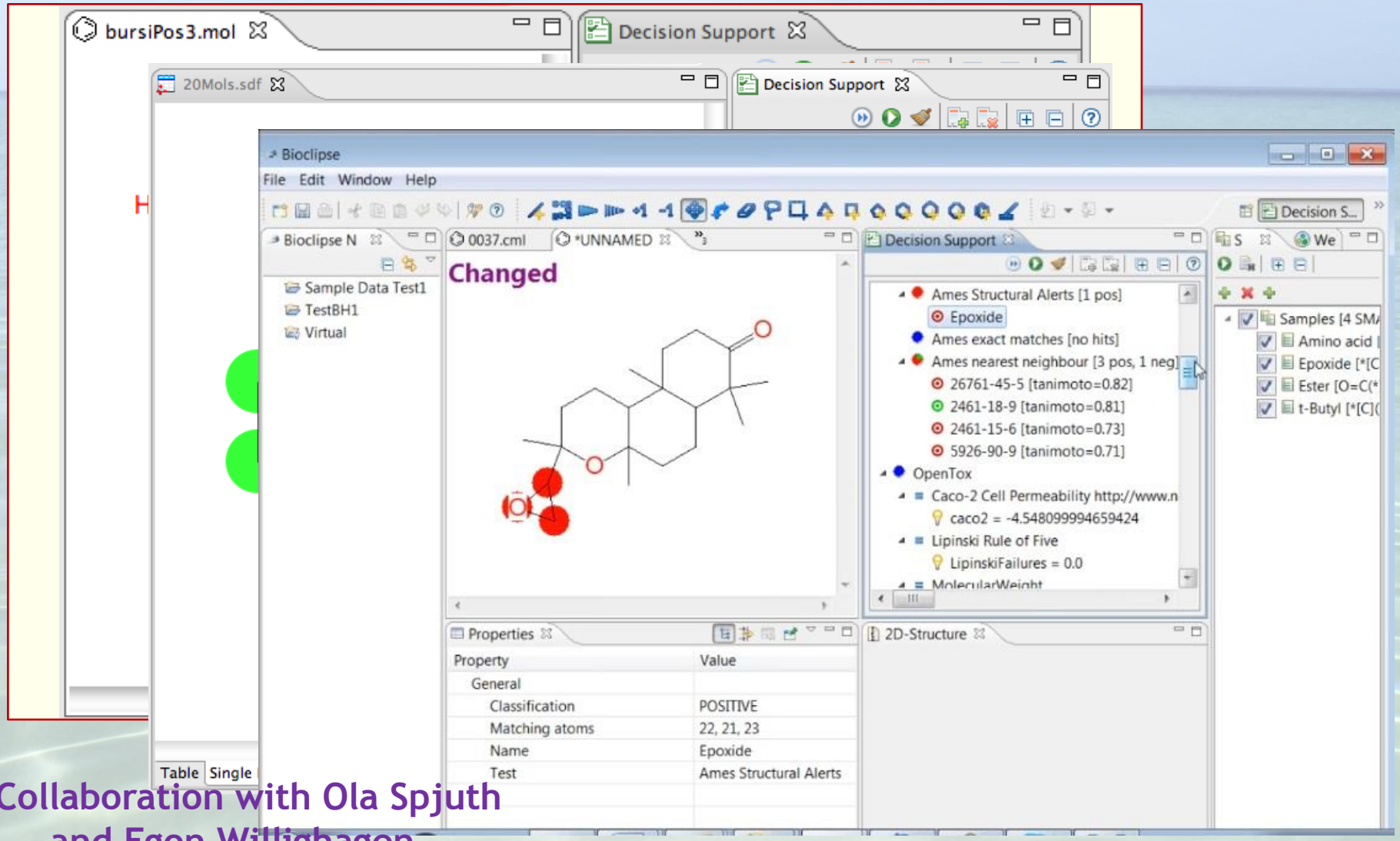
	OECD Principle	OpenTox addresses Validation Principles by...
1	**Defined Endpoint**	providing a unified source of well defined and documented toxicity data with a common vocabulary
2	**Unambiguous Algorithm**	providing transparent access to well documented models and algorithms as well as to the source 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

Integrating public and confidential data



Use Open Standards on Resources but with extensive Authorisation and Authentication facilities accompanied by confidential data policies. e.g. *Validation against Confidential Data Case* implemented by OpenTox Spring 2011

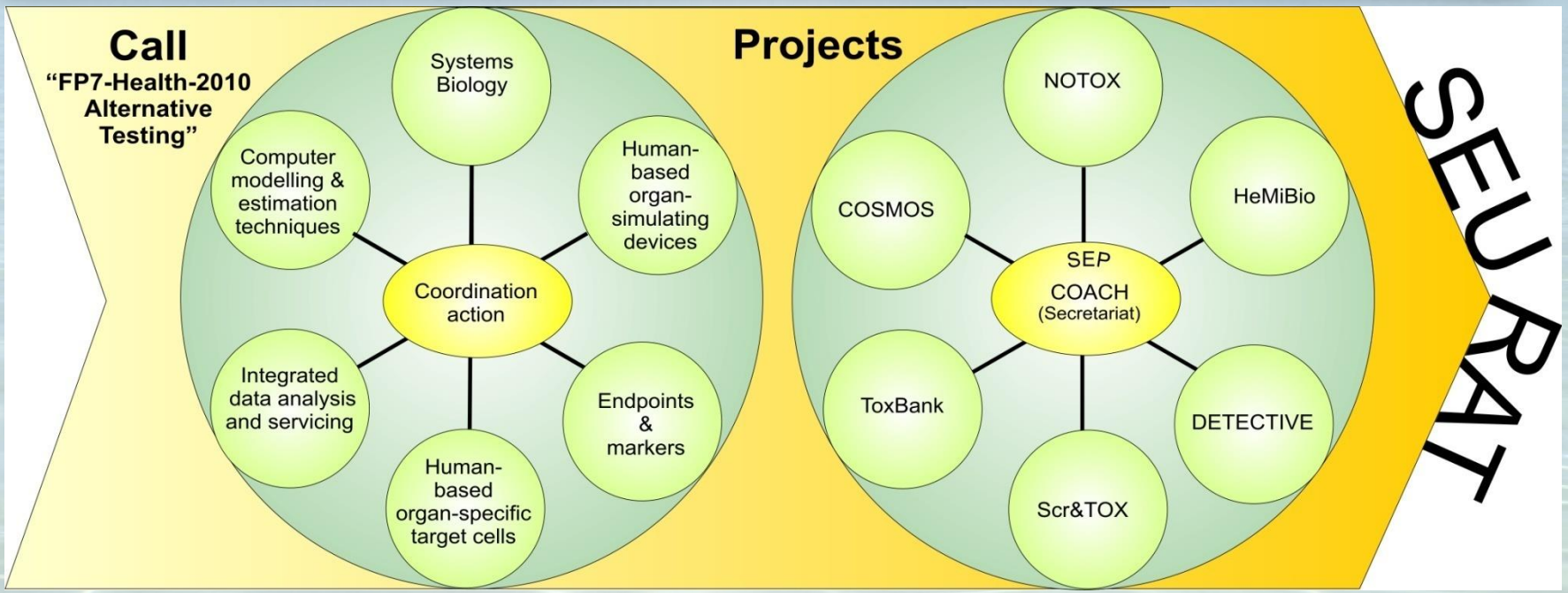
Bioclipse Visualisation Workbench - OpenTox



Collaboration with Ola Spjuth and Egon Willighagen

Bioclipse-OpenTox Integration – See Application example in Chapter in **Open Source Software in Life Science Research: Practical Solutions to Common Challenges in the Pharmaceutical Industry and Beyond (Woodhead Publishing Series in Biomedicine) edited by Lee Harland and Mark Forster (30 Oct 2012)**

The Building Blocks of SEURAT-1



~ 70 research groups from European Universities, Public Research Institutes and Companies (more than 30% SMEs)

www.seurat-1.eu

ToxBank

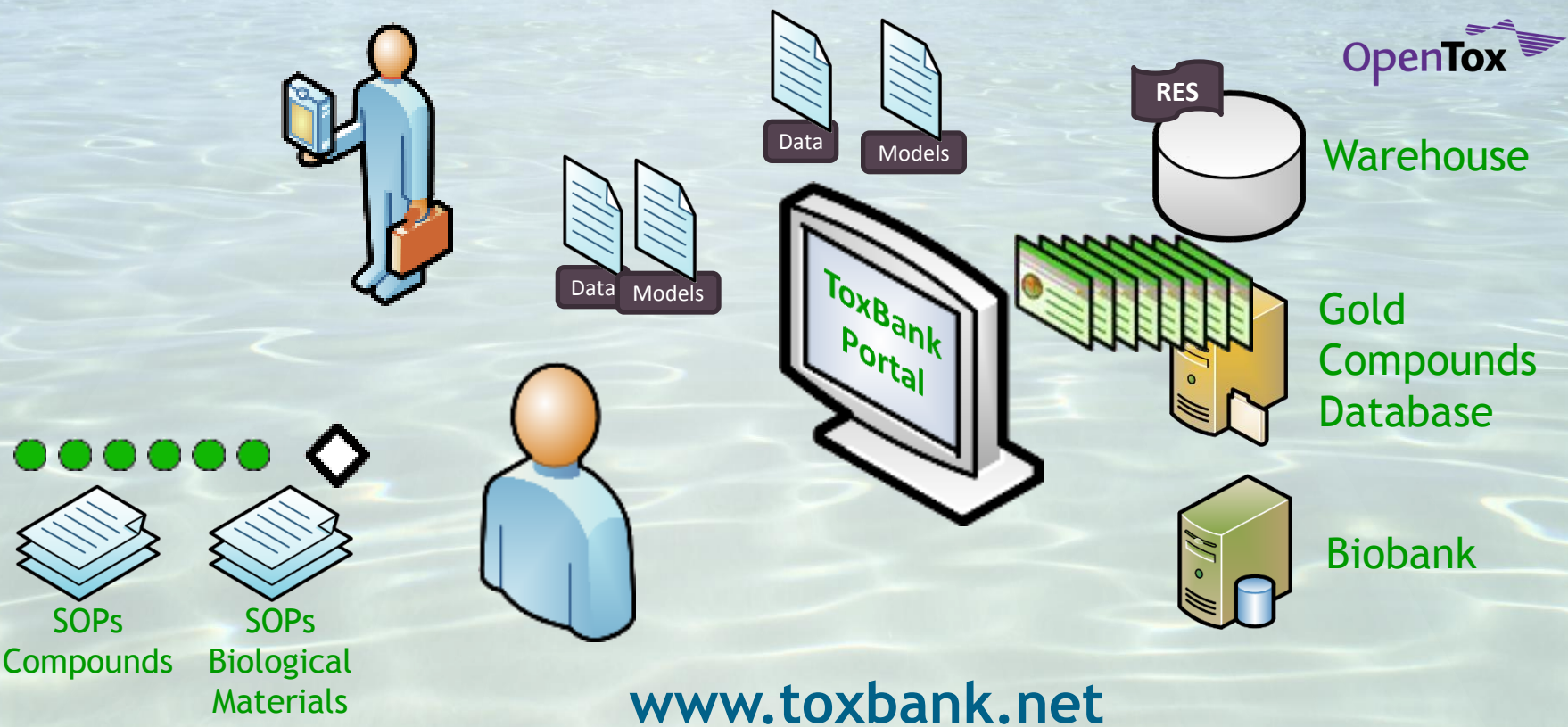
This project is jointly funded by Cosmetics Europe and the EC. Any opinions expressed in this slide are those of the author. Cosmetics Europe is not liable for any use that may be made of the information contained therein.

Cosmetics Europe
the personal care association

SEVENTH FRAMEWORK PROGRAMME

Our Infrastructure Vision for ToxBank supporting all steps of Predictive Toxicology Research

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



www.toxbank.net

ToxBank

Cosmetics Europe

SEVENTH FRAMEWORK PROGRAMME

Working on Requirements & Systems Design

Organizing notes collected from interviews with SEURAT scientists

Hierarchically grouping the notes

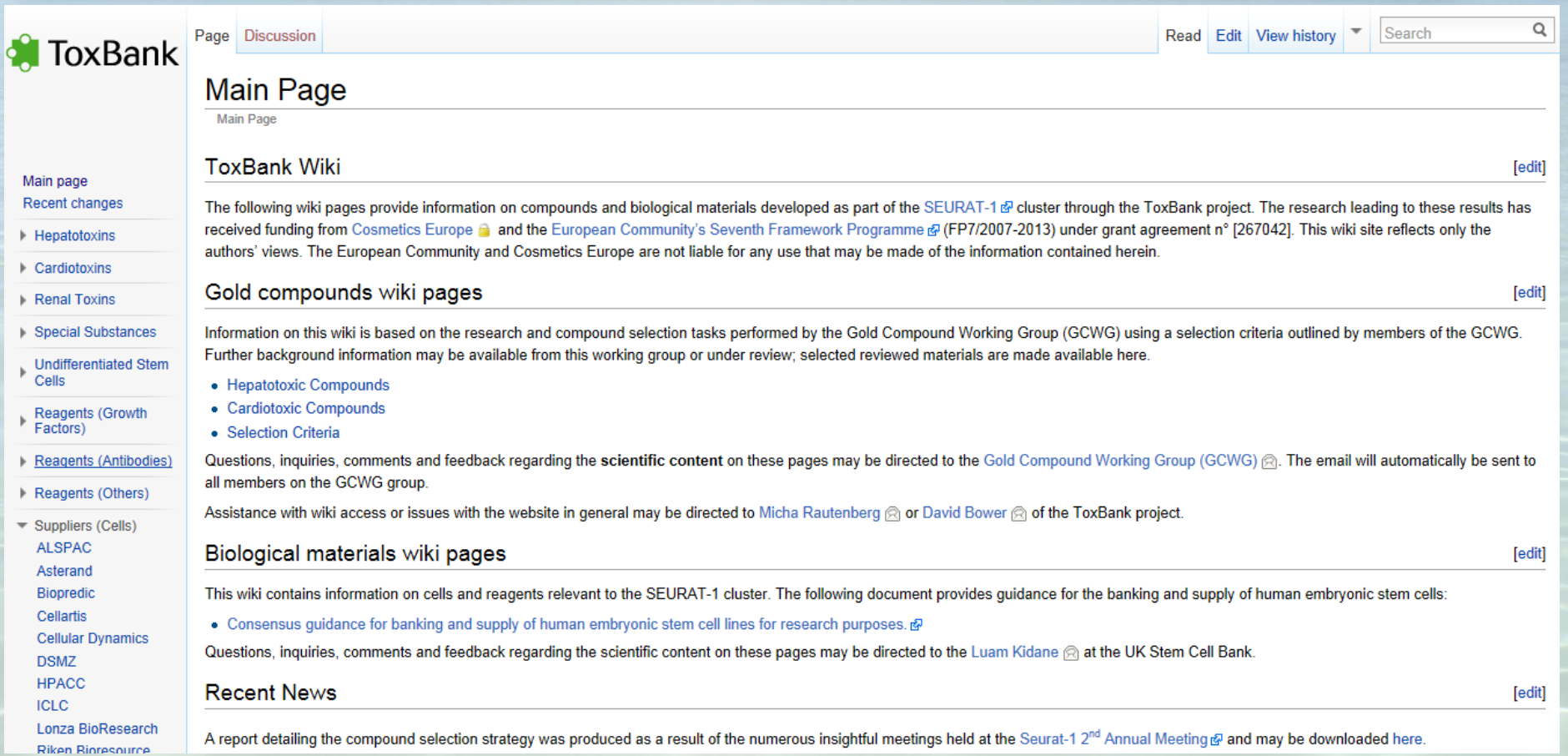
Generating design ideas

Reviewing use cases

Evaluating different

Storyboarding different user interfaces

ToxBank Wiki Development



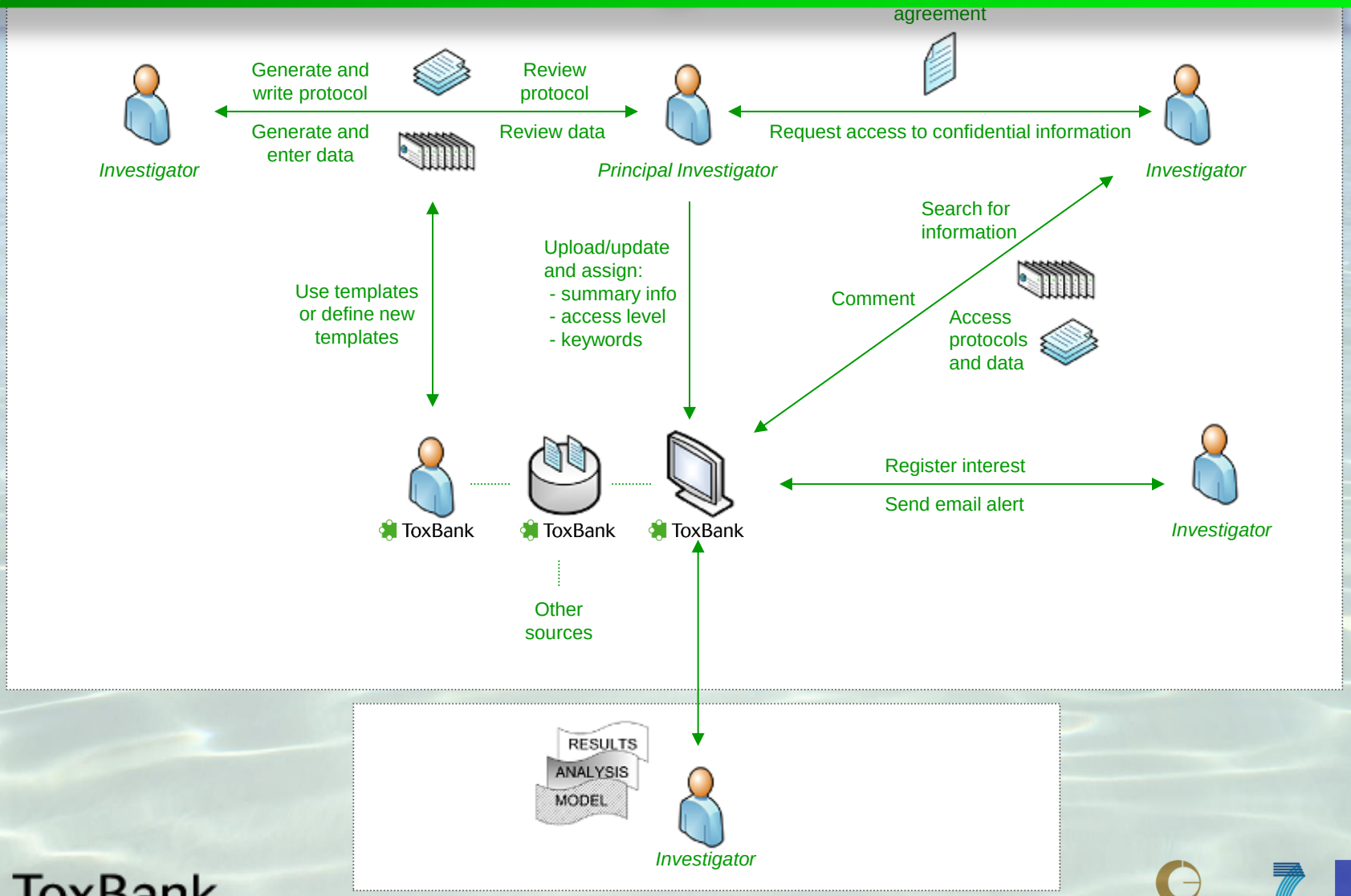
ToxBank

wiki.toxbank.net

Cosmetics Europe

SEVENTH FRAMEWORK PROGRAMME

Data Sharing Processes



ToxBank

Phase 2: Integrated data analysis

Cosmetics Europe

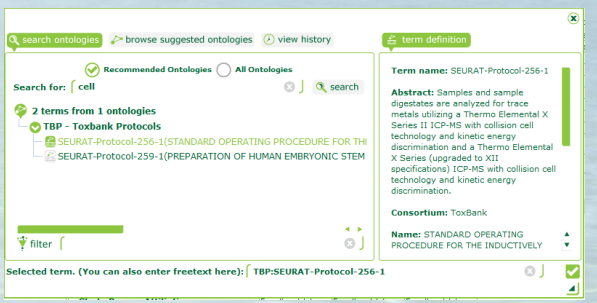
SEVENTH FRAMEWORK PROGRAMME

Unified data access

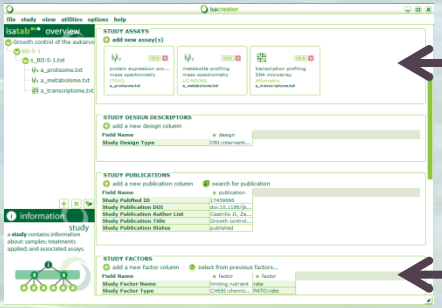
Investigation information

SEURAT-1 information

Publications

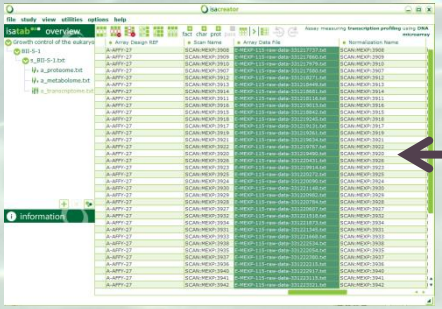


Each step linked to a SEURAT-1 protocol

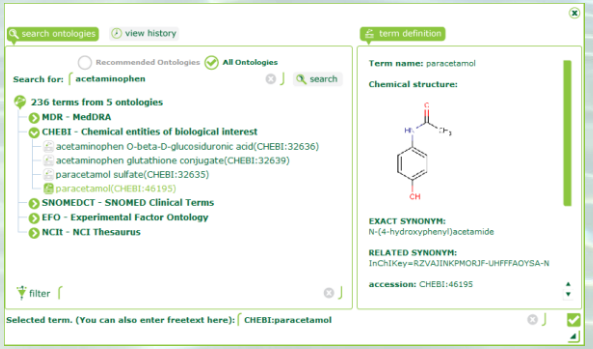


Templates for different assays

Specify experimental factors



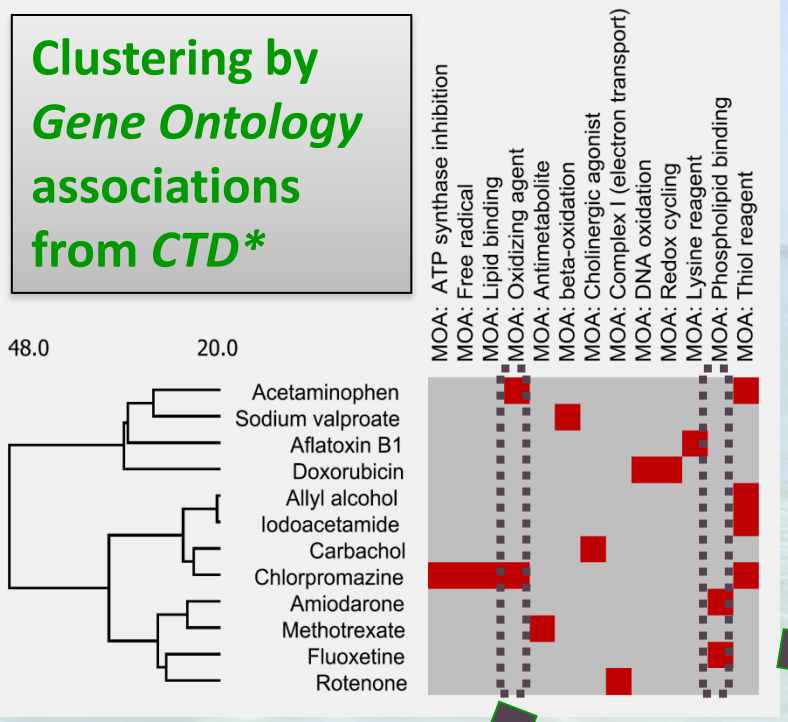
Materials and results, with links to files containing the raw or processed data



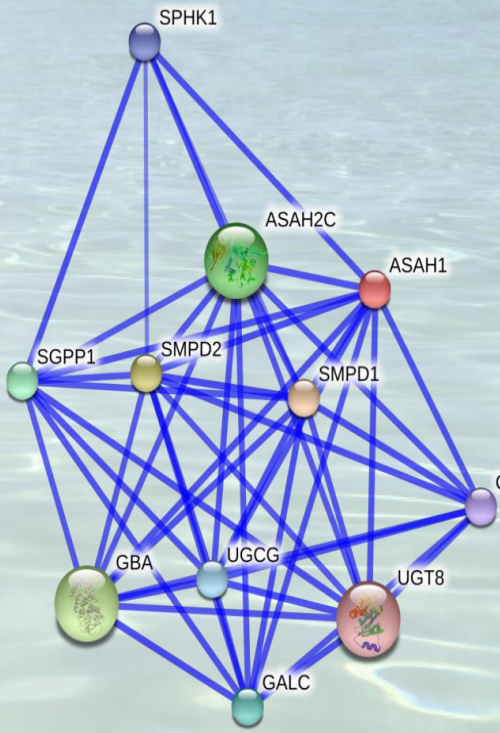
Terms mapped to ontologies

Public Data Analysis

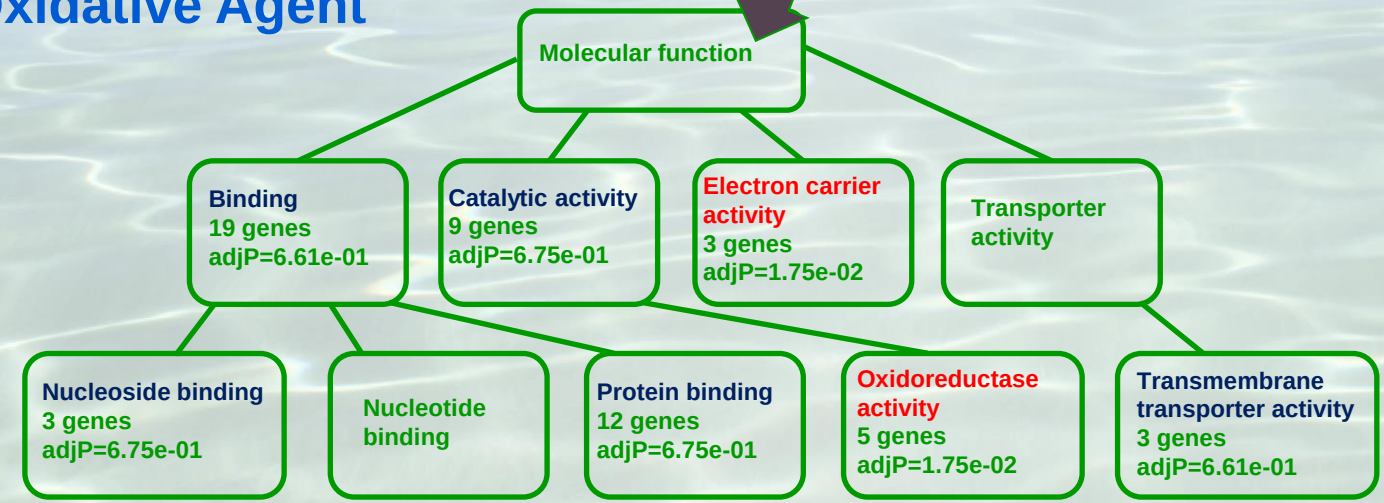
Clustering by *Gene Ontology* associations from *CTD**



Phospholipid Binding



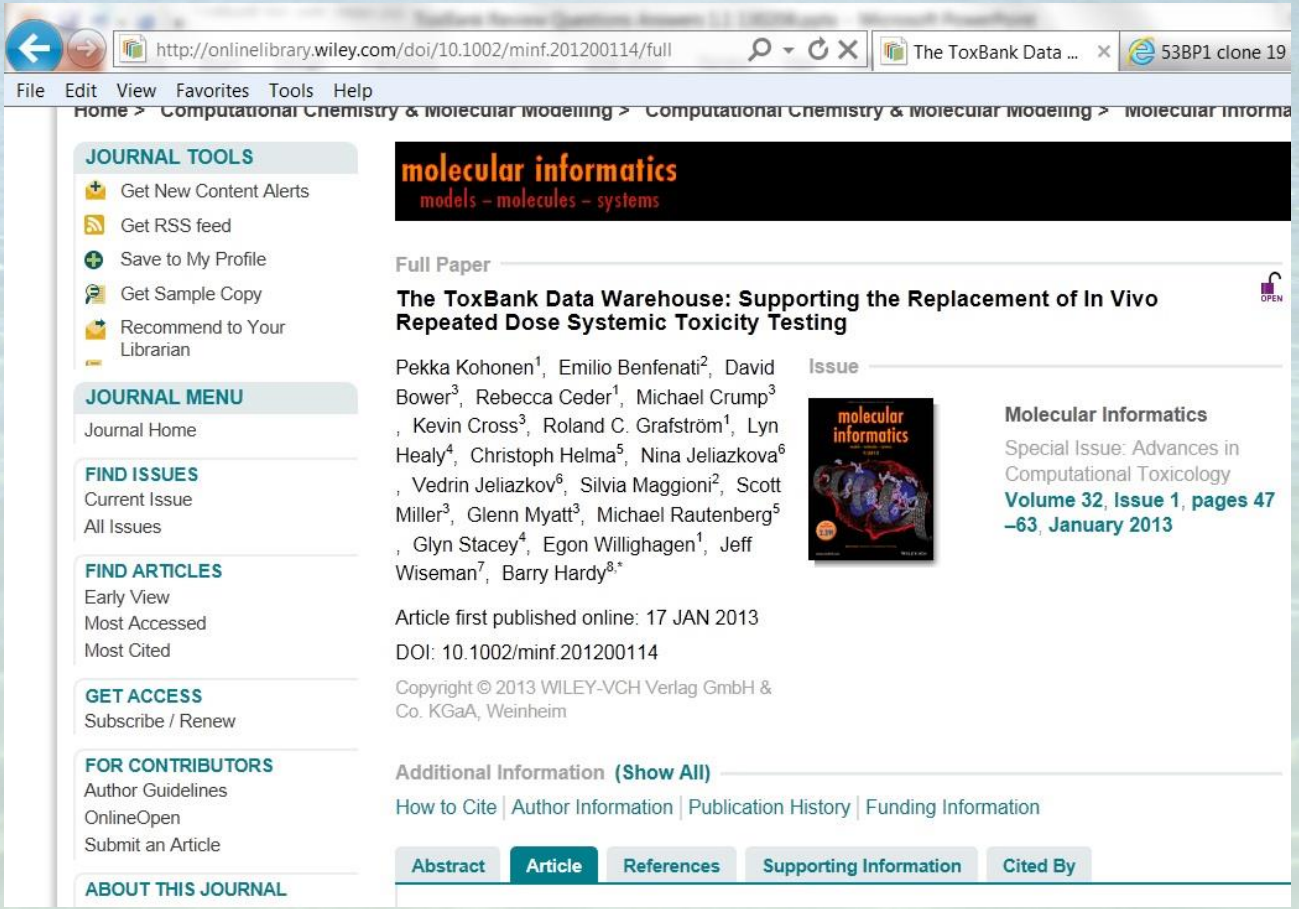
Oxidative Agent



**CTD = Comparative Toxicogenomics Database (www.ctd.org)*

Kohonen P. et al. The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing. Mol. Inf.17 JAN 2013.

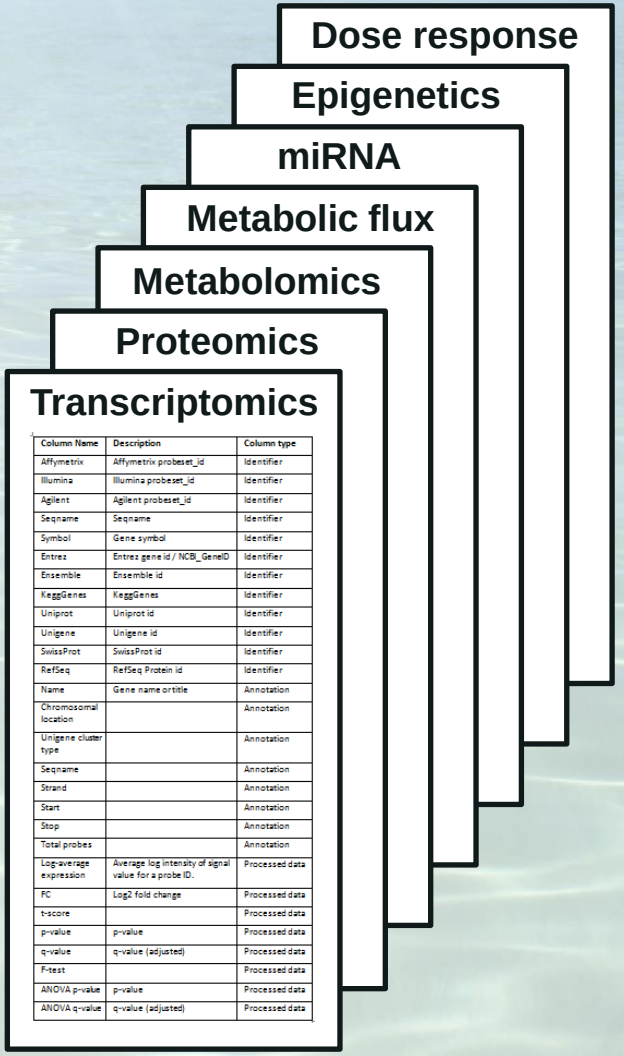
ToxBank Phase I – Unified data access



onlinelibrary.wiley.com/doi/10.1002/minf.201200114/full

Standardization of processed data

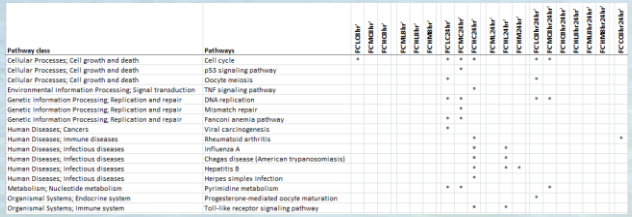
- To support ToxBank integrated data analysis objectives (precise searching, meta analysis, …)
- The columns
  - (1) uniquely identify the *material* (e.g. the Affymetrix probeset_id),
  - (2) annotate the *material* (e.g. the name of the gene),
  - (3) describe the processed results (e.g. fold change comparing genes expressed in the treated sample to the control).



Column Name	Description	Column type
Affymetrix	Affymetrix probeset_id	Identifier
Illumina	Illumina probeset_id	Identifier
Agilent	Agilent probeset_id	Identifier
Seqname	Seqname	Identifier
Symbol	Gene symbol	Identifier
Entrez	Entrez gene id / NCBI_GeneID	Identifier
Ensemble	Ensemble id	Identifier
KeggGenes	KeggGenes	Identifier
Uniprot	Uniprot id	Identifier
Unigene	Unigene id	Identifier
SwissProt	SwissProt id	Identifier
RefSeq	RefSeq Protein id	Identifier
Name	Gene name ortitle	Annotation
Chromosomal location		Annotation
Unigene cluster type		Annotation
Seqname		Annotation
Strand		Annotation
Start		Annotation
Stop		Annotation
Total probes		Annotation
Log-average expression	Average log intensity of signal value for a probe ID.	Processed data
FC	Log2 fold change	Processed data
t-score		Processed data
p-value	p-value	Processed data
q-value	q-value (adjusted)	Processed data
F-test		Processed data
ANOVA p-value	p-value	Processed data
ANOVA q-value	q-value (adjusted)	Processed data

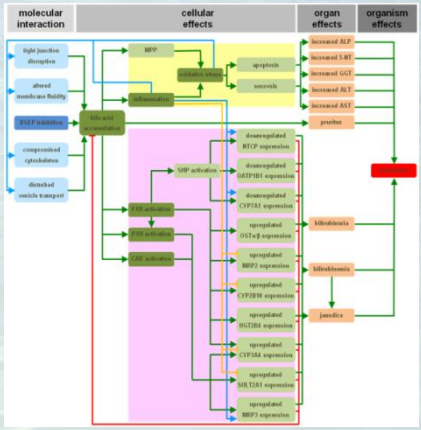
Analysis examples

Multi-omics pathway enrichment

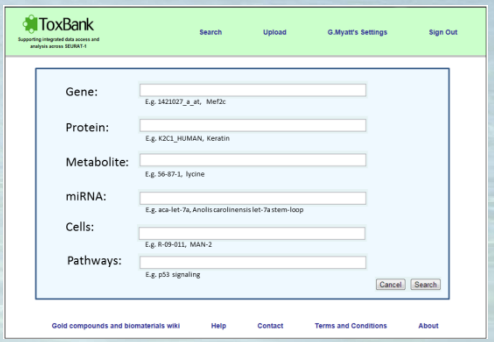


Pathway class	Pathways
Cellular Processes; Cell growth and death	Cell cycle
Cellular Processes; Cell growth and death	p53 signaling pathway
Cellular Processes; Cell growth and death	Oocyte meiosis
Environmental Information Processing; Signal transduction	TNF signaling pathway
Genetic Information Processing; Replication and repair	DNA replication
Genetic Information Processing; Replication and repair	Mismatch repair
Genetic Information Processing; Replication and repair	Fanconi anemia pathway
Human Diseases; Cancers	Viral carcinogenesis
Human Diseases; Immune diseases	Rheumatoid arthritis
Human Diseases; Infectious diseases	Influenza A
Human Diseases; Infectious diseases	Chagas disease (American trypanosomiasis)
Human Diseases; Infectious diseases	Hepatitis B
Human Diseases; Infectious diseases	Herpes simplex infection
Metabolism; Nucleotide metabolism	Pyrimidine metabolism
Organismal Systems; Endocrine system	Progesterone-mediated oocyte maturation
Organismal Systems; Immune system	Toll-like receptor signaling pathway

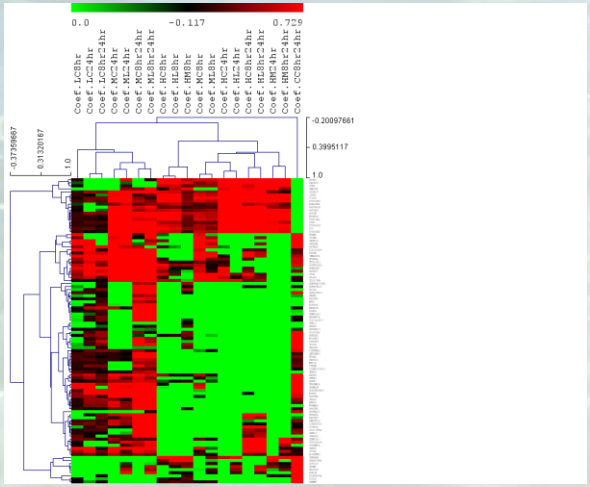
Development of AOPs



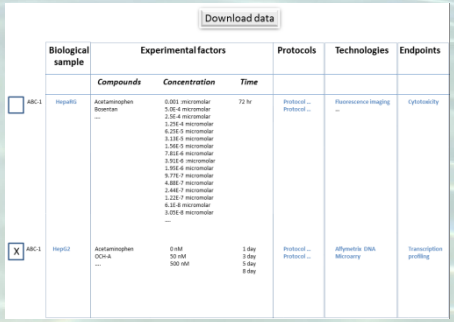
Search other investigations



Analysis and visualization



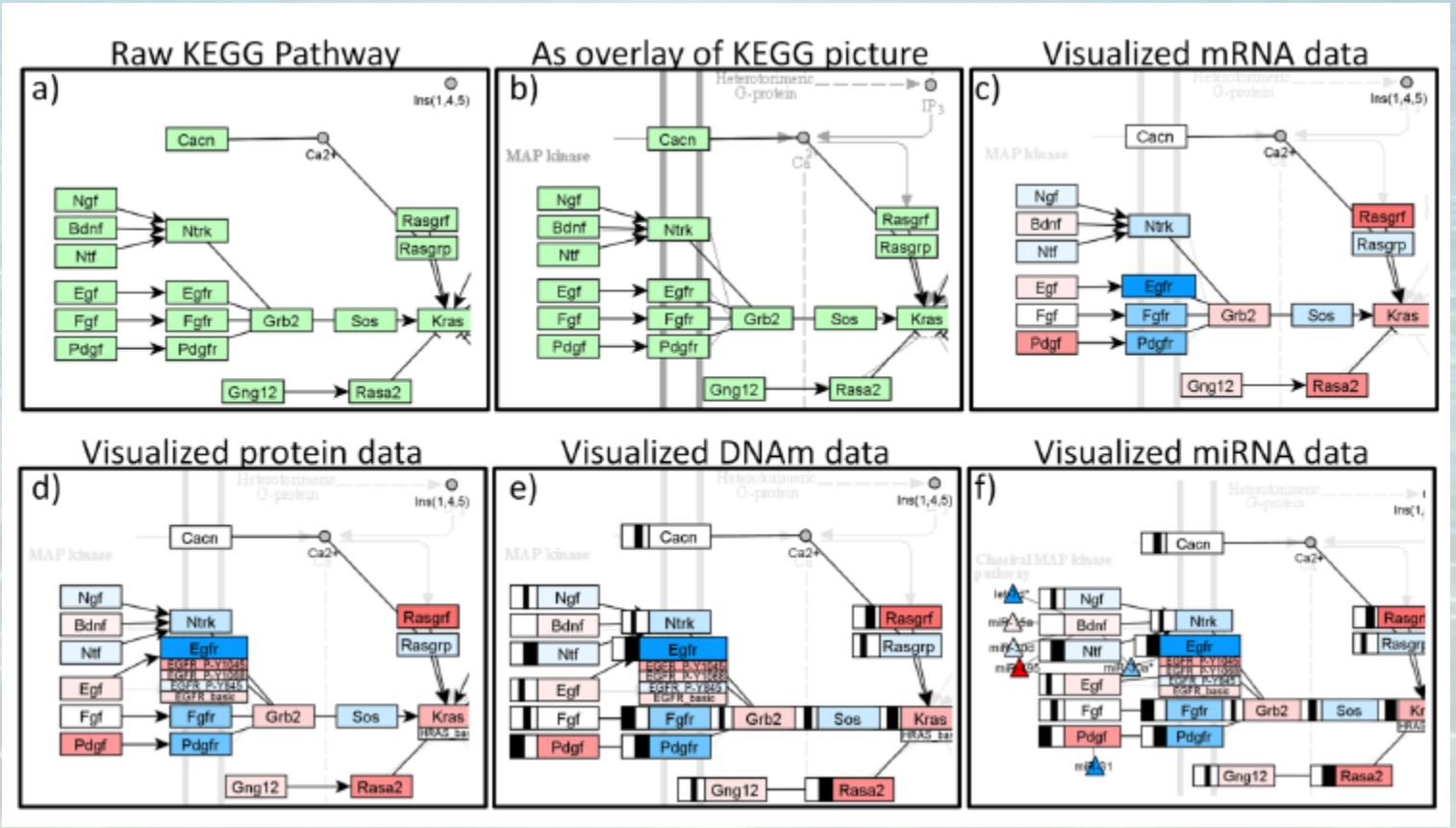
Understanding experiments



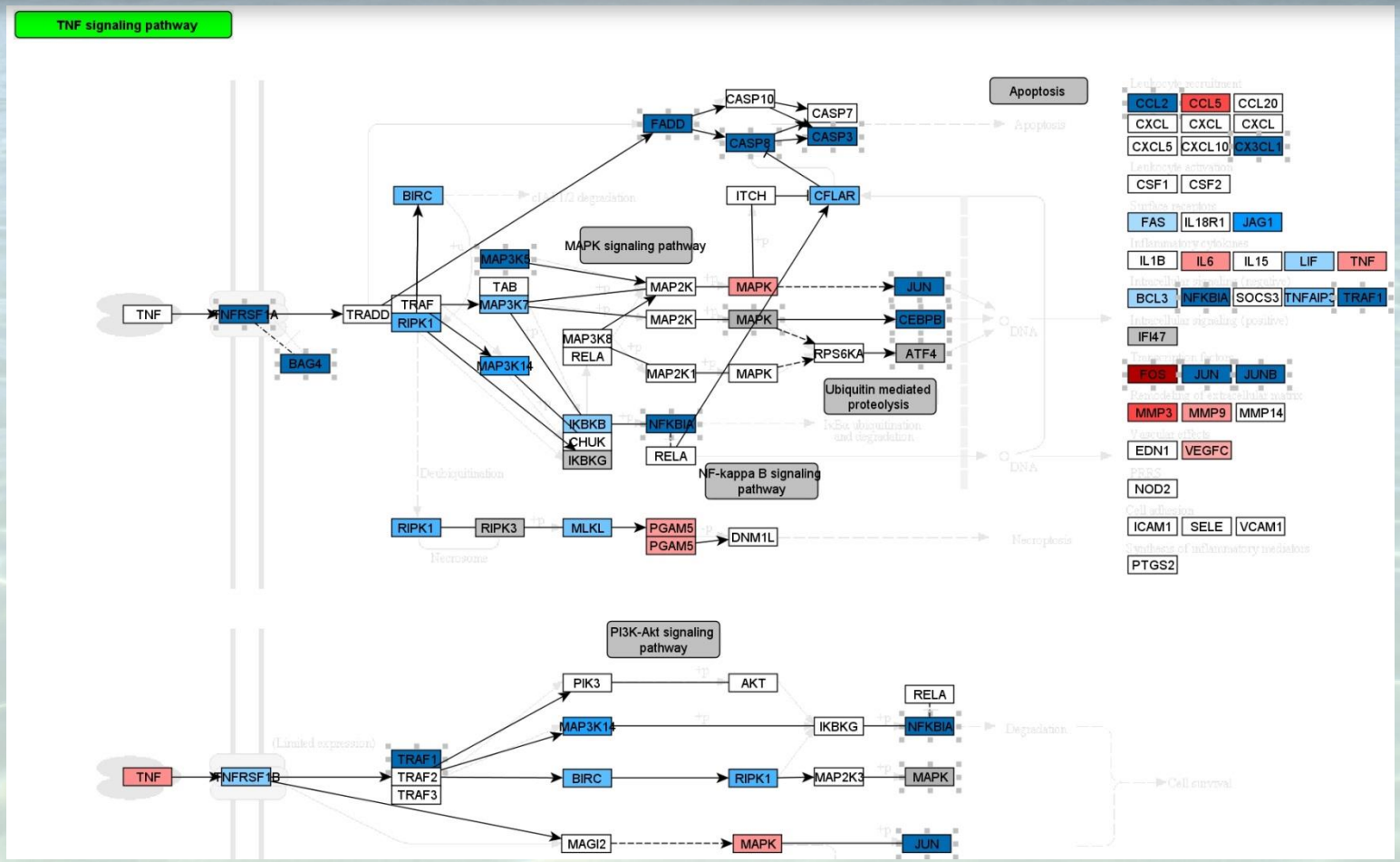
Understanding kinetics

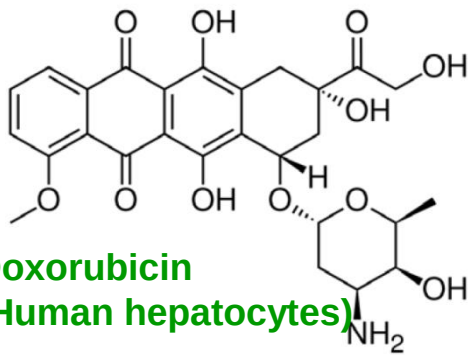
PK-ADME [7]	Compound Assessment
PK parameters [7]	The most commonly used dose schedule when used as a single agent is 60 to 75 mg/m2 as a single intravenous injection administered at 21-day intervals.
	Protein binding 70%
	Half life 55 hours
	V_d 20-30 L/kg (700-100 L/m^2)
	C_{max} 3 uM for 30 mg/m^2 intravenous bolus dose. Cellular levels are about 30–100-fold higher than that of the plasma.
	Excretion predominantly in bile, 40-50% in feces within 7 days (50% as unchanged drug).
	Plasma clearance 324 to 809 mL/min/m^2, biphasic
	Metabolism ~50% metabolized by the liver
	References:
	- http://www.drugbank.ca/drugs/DB00997
	- http://reference.medscape.com/drug/doxorubicin-342120
	- A.-K. Souid et al. "Immediate effects of anticancer drugs on mitochondrial oxygen consumption", Biochemical Pharmacology 66 (2003) 977–987

Understanding multiple omics datasets



ToxBank – TG Gates integrated analysis example



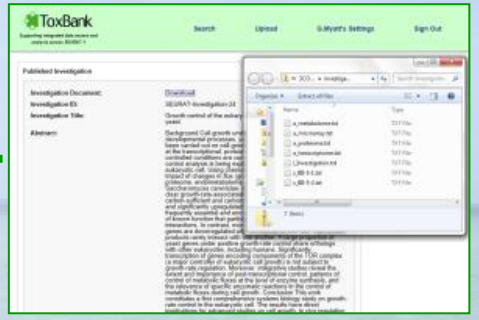


Doxorubicin (Human hepatocytes)

Transcriptomics profiles

A

ToxBank Data Warehouse (data curation and retrieval)



Protocols and SOPs, upload investigation data in ISA-TAB format

C

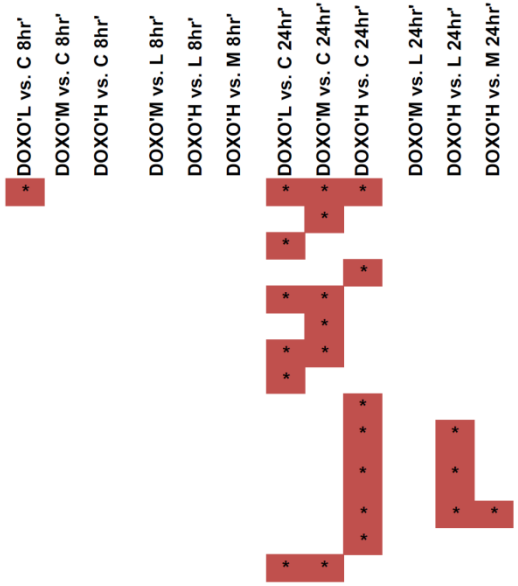
Differentially expressed genes (R/Bioconductor)

B

Pathway meta-analysis using KEGG pathways (InCroMap) software)

Pathways

1. Cell cycle
2. p53 signaling pathway
3. Oocyte meiosis
4. TNF signaling pathway
5. DNA replication
6. Mismatch repair
7. Fanconi anemia pathway
8. Viral carcinogenesis
9. Rheumatoid arthritis
10. Influenza A
11. Chagas disease (American trypanosomiasis)
12. Hepatitis B
13. Herpes simplex infection
14. Pyrimidine metabolism



Significance: *=FDR q-value < 0.05

Doses: C=Control, L=Low, M=Middle, H=High; Time: 8hr=8 hours, 24hr=24 hours

Connectivity Map (MCF7, PC-3 cell lines; p < 0.01)

1. Doxorubicin (0.999) *
2. H-7 (0.999) *
3. Mitoxantrone (0.998) *
4. Alsterpaullone (0.997) *
5. Camptothecin (0.991)
6. Ronidazole (0.87)
7. Medrysone (0.817)
8. Gliclazide (0.777)
9. Ginkgolide A (0.776)
10. Ellipticine (0.746) *
11. Etamsylate (0.746)
12. Trioxysalen (0.744)
13. Ethaverine (0.739)
14. Doxazosin (0.738)
15. Amiodarone (0.719)
16. Morantel (0.687)
17. Phthalylsulfathiazole (0.684)
18. Dipyridamole (0.672)
19. Demeclocycline (0.645)
20. Famprofazone (0.643)

*= topoisomerase II inhibitor (Mantra 2.0)

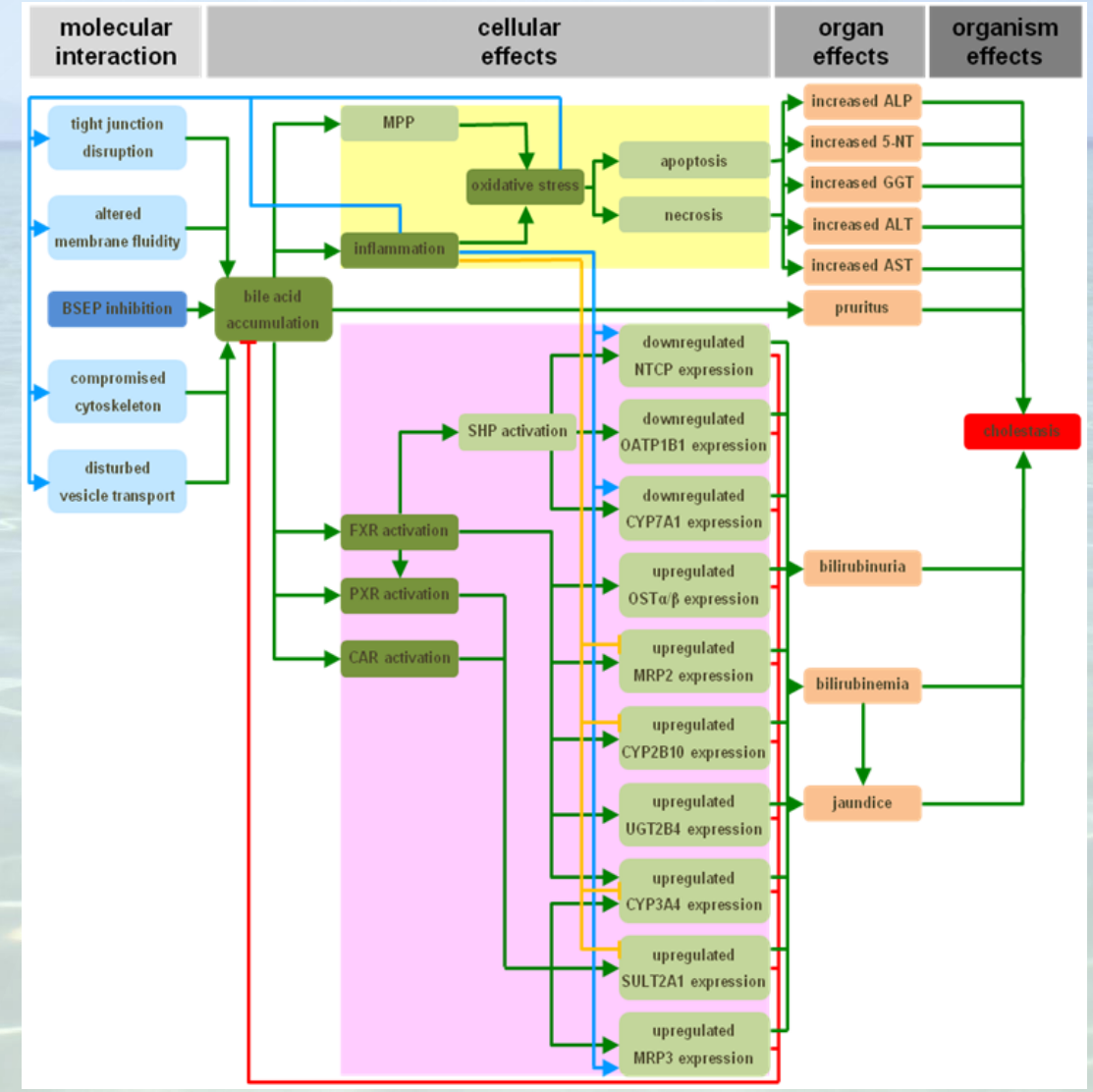
D

Comparative Toxicogenomics Database (q-value < 0.01)

Disease Name	Disease ID
1. Cardiovascular Diseases	MESH:D002318
2. Digestive System Diseases	MESH:D004066
3. Neoplasms	MESH:D009369
4. Neoplasms by Histologic Type	MESH:D009370

Kohonen P, Ceder R, Smit I, Hongisto V, Myatt G, Hardy B, Spjuth O, Grafström R. Basic Clin Pharmacol Toxicol. 2014 Jul;115(1):50-8.

Adverse outcome pathway (AOP) : drug-induced cholestasis



Vinken M., Landesmann B., Goumenou M., Vinken S., Shah I., Jaeschke H., Willett C., Whelan M., Rogiers V. (2013) Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. *Archives of Toxicology*: submitted .

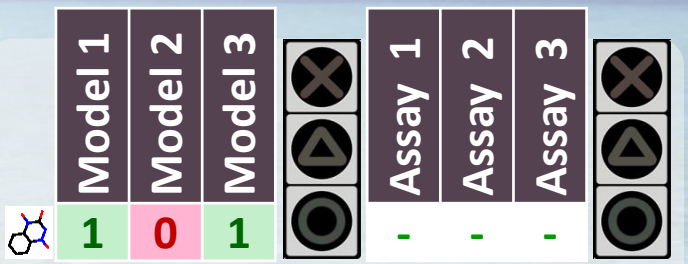
Event Driven Weight of Evidence

collaboration MODERATOR

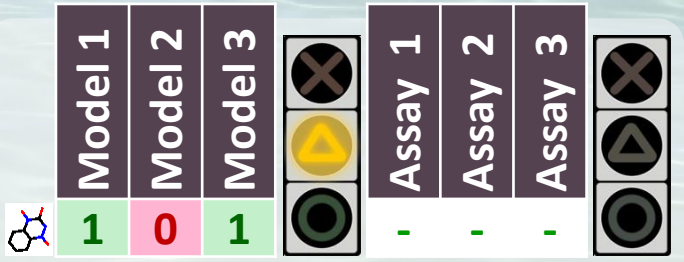
Consensus Rule Editor

Recommendation Rules:

Drug Discovery Dashboard



Synergy

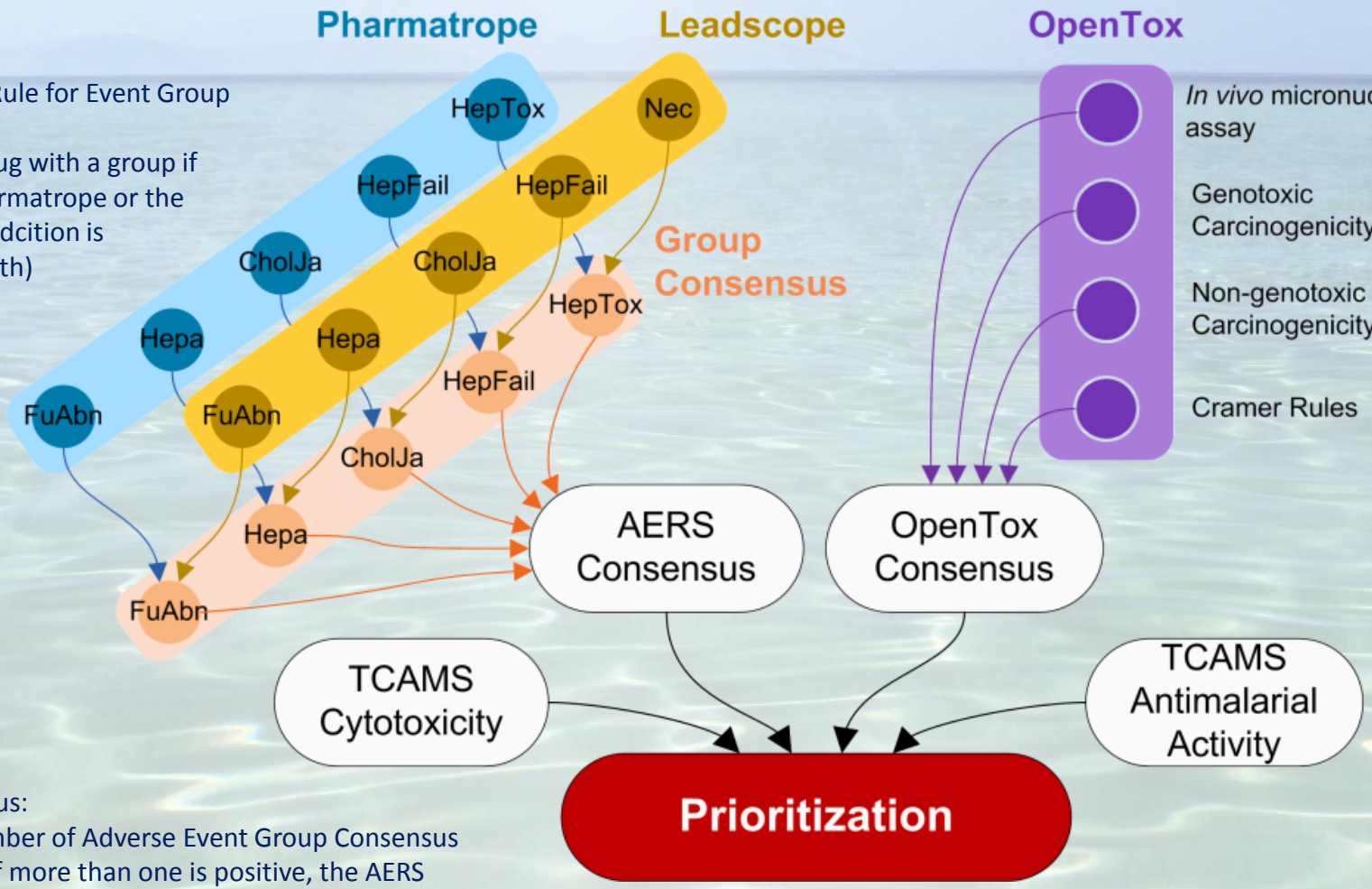


OpenTox

Application to Anti-Malarial Drug Discovery – www.scientistsagainstmalaria.net

Combining Predictions and Experimental Data

Combination Rule for Event Group Predictions:
Associate a drug with a group if either the Pharmatrope or the Leadscope predcition is positive (or both)



AERS Consensus:
Count the number of Adverse Event Group Consensus associations. If more than one is positive, the AERS Consensus is positive.

OpenTox Consensus:
Negative if both carcinogenicity and the micronucleus assay predictions are negative, OR if the Cramer Rule classification is Class I. Positive otherwise.

TCAMS Cytotoxicity:
Positive if > 30% growth inhibition at 10 µM.

TCAMS Antimalarial Activity:
Positive if > 80% growth inhibition of P. Falciparum DD2 at 2 µM.

Main objectives of eNanoMapper

- Modular infrastructure for data storage, sharing and searching, based on open standards and semantic web technologies, minimum information standards and established security solutions;
- Development of ontologies for the categorisation and characterisation of eNMs in collaboration with other projects
- Creation of new computational models in nanomaterials safety through the implementation of interfaces for toxicity modelling and prediction algorithms which may process all data made available through eNanoMapper (e.g. using algorithms available from the OpenTox FP7 project or statistical/data mining software)
- Meta analysis of nano-bio interactions supporting "safe-by-design" ENMs development by pursuing a Linked Data approach which integrates data and metadata originating from diverse sources within nanoscience, chemistry, biology and toxicology
- Creation of tools for the exchange, quality assurance and reporting of research protocols and data for regulatory purposes
- Creation of a community framework for interdisciplinary collaboration

www.enanomapper.net

NanoEHS Data & Modeling & EU-US Cooperation – Themes of our Discussion (Venice, March 2015)

Reference Information

More Open Data

Interoperability

Transparency

Ontology

Sustainability

Integrated

Use cases

Reuse

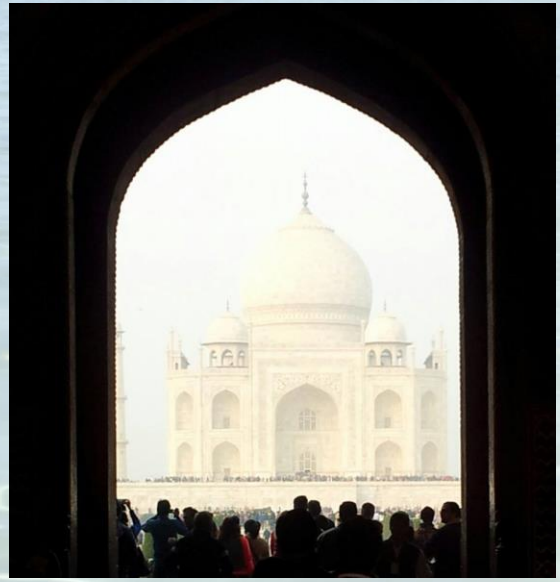
Open Source

Knowledge

Systems Design

Create Solutions

Taj Mahal Project Action Plan – Supporting EU-US Community of Research InterAction



1. Define Use Cases (with all CoRS)
2. Create Systems Design
3. Evaluate Existing Resources
4. Develop Sustainability Plan
5. Create Application based on Open data, software, protocols, and ontology (Agile Development)
6. Support collaborative work of all CoRs with using the Application

OpenTox Association

1) International Non-profit Association founded 27 March 2015
2) Individual Memberships
3) Working Groups – prioritised activities and goals
4) Meetings

a. 30 Sept – 2 Oct, Dublin, Ireland
b. 13 – 15 Oct, Philadelphia, USA
c. 7 – 9 Dec, Seoul, South Korea

OpenTox Working Groups

Working Groups

a. Application Programming Interfaces (APIs), *Christoph Helma (in silico toxicology)*
b. Data, Metadata and Ontology Standards, *Thomas Exner (Douglas Connect GmbH)*
c. Adverse Outcome Pathway (AOP) development, *Stephen Edwards (US EPA)*
d. Deployment, *Tim Dudgeon (Informatics Matters)*

Further information under www.opentox.net and www.opentox.org (content currently being reviewed, updated and merged)

Collaborating Partners on eNanoMapper

Douglas Connect, Switzerland (Coordinator)

Maastricht University, Netherlands

In Silico Toxicology, Switzerland

National Technical University of Athens, Greece

Ideaconsult, Bulgaria

EMBL-EBI, UK

Karolinska Instituet, Sweden

VTT, Finland

Associate Partners

Collaborating Partners on OpenTox

Douglas Connect, Switzerland (Coordinator)

In Silico Toxicology, Switzerland

Albert Ludwigs University Freiburg, Germany

Ideaconsult, Bulgaria

National Technical University of Athens, Greece

Istituto Superiore di Sanità, Italy

Fraunhofer Institute for Toxicology & Experimental Medicine, Germany

Technical University of Munich, Germany

David Gallagher, UK

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

Seascape Learning & JNU, India

ToxBank Acknowledgements

UK Stem Cell Bank, NIBSC-HPA

Ideaconsult Ltd