

OpenTox framework for predictive toxicology

Семинар-работна среща по Биоматематика

May 26-27, 2010

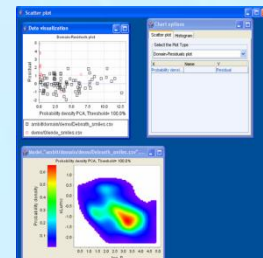
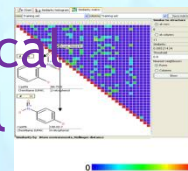
Др. Нина Желязкова
IdeaConsult



Introduction

- Develops and maintains several well-known open source software products, in particular
 - Toxtree (used in the application of the threshold of toxicological concern)
 - Toxmatch (for encoding and applying chemical similarity indices)
 - Ambit (a QSAR decision support system, including generic database management, structure conversions and searching, as well as applicability domain assessment).
- These tools have been carefully designed and developed in close co-operation both with academia, regulatory bodies and industry.

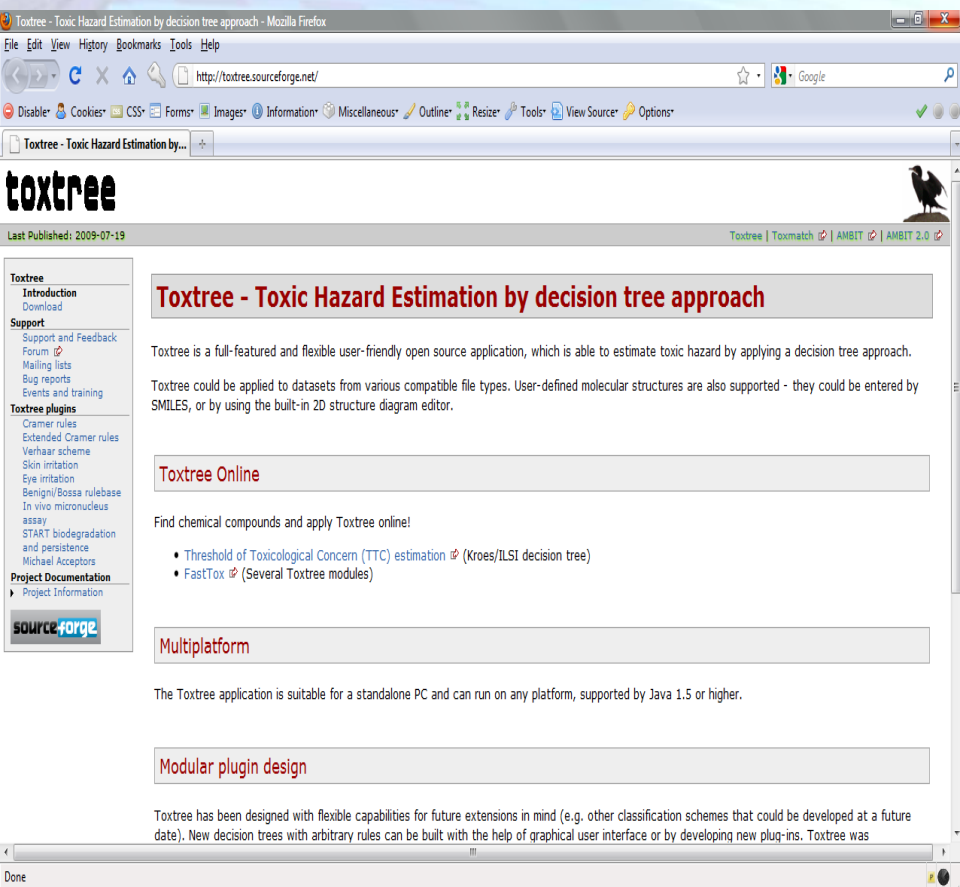
- Toxtree 1.60 - estimates toxic hazard by applying a decision tree approach
- Toxmatch 1.06 - A chemical similarity evaluation tool
- Ambit Discovery
- Ambit Database Tools
- QMRF repository
- Ambit XT
- Partner in OpenTox FP7 project
- Partner in CADASTER FP7 project



Toxtree 1.60

<http://toxtree.sourceforge.net>

Applied in industry and academy worldwide



The screenshot shows the Toxtree website in a Mozilla Firefox browser. The address bar displays <http://toxtree.sourceforge.net/>. The website features the Toxtree logo, a navigation menu, and a main content area with the following sections:

- Toxtree - Toxic Hazard Estimation by decision tree approach**

Toxtree is a full-featured and flexible user-friendly open source application, which is able to estimate toxic hazard by applying a decision tree approach.

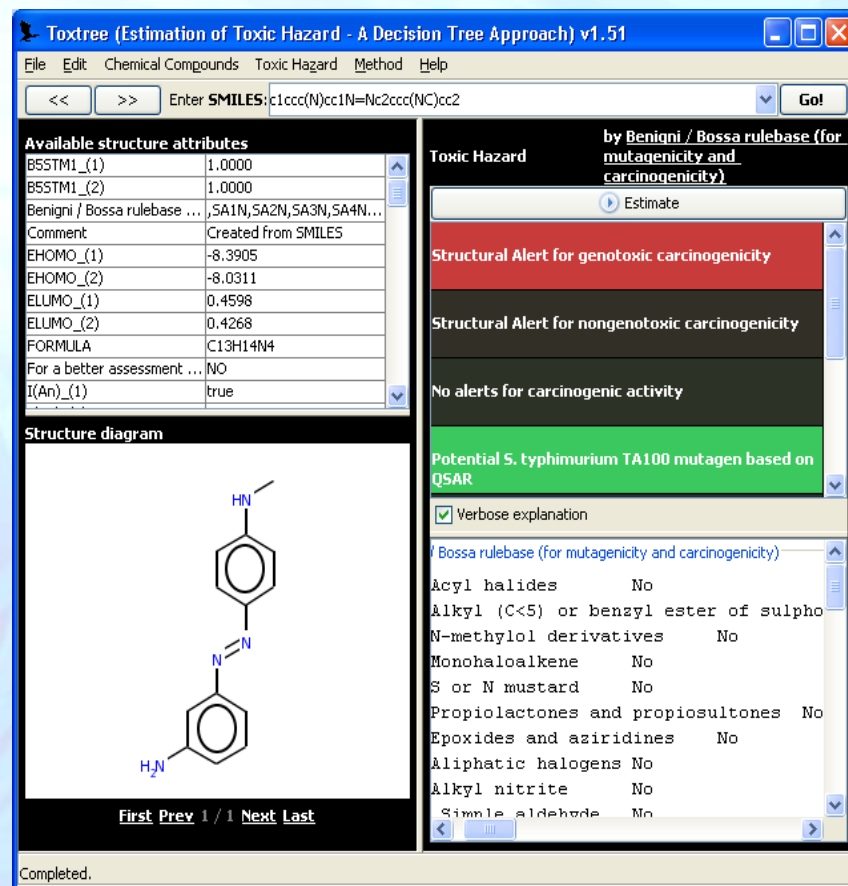
Toxtree could be applied to datasets from various compatible file types. User-defined molecular structures are also supported - they could be entered by SMILES, or by using the built-in 2D structure diagram editor.
- Toxtree Online**

Find chemical compounds and apply Toxtree online!

 - [Threshold of Toxicological Concern \(TTC\) estimation](#) (Kroes/ILSI decision tree)
 - [FastTox](#) (Several Toxtree modules)
- Multipatform**

The Toxtree application is suitable for a standalone PC and can run on any platform, supported by Java 1.5 or higher.
- Modular plugin design**

Toxtree has been designed with flexible capabilities for future extensions in mind (e.g. other classification schemes that could be developed at a future date). New decision trees with arbitrary rules can be built with the help of graphical user interface or by developing new plug-ins. Toxtree was



The screenshot shows the Toxtree application window titled "Toxtree (Estimation of Toxic Hazard - A Decision Tree Approach) v1.51". The interface includes a menu bar, a toolbar, and several panels:

- Enter SMILES:** c1ccc(N)cc1N=Nc2ccc(NC)cc2
- Available structure attributes:**

Attribute	Value
BSSTM1_(1)	1.0000
BSSTM1_(2)	1.0000
Benigni / Bossa rulebase ...	,SA1N,SA2N,SA3N,SA4N...
Comment	Created from SMILES
EHOMO_(1)	-8.3905
EHOMO_(2)	-8.0311
ELUMO_(1)	0.4598
ELUMO_(2)	0.4268
FORMULA	C13H14N4
For a better assessment ...	NO
I(An)_(1)	true
- Structure diagram:** A 2D chemical structure diagram of the compound, showing a benzene ring with an amino group (-NH2) and a diazo group (-N=N-) attached to another benzene ring with an amino group (-NH2).
- Toxic Hazard:**
 - by Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)
 - Estimate
 - Structural Alert for genotoxic carcinogenicity
 - Structural Alert for nongenotoxic carcinogenicity
 - No alerts for carcinogenic activity
 - Potential *S. typhimurium* TA100 mutagen based on QSAR
 - ☒ Verbose explanation
- / Bossa rulebase (for mutagenicity and carcinogenicity):**

Rule	Result
Acyl halides	No
Alkyl (C<5) or benzyl ester of sulpho	No
N-methylol derivatives	No
Monohaloalkene	No
S or N mustard	No
Propiolactones and propiosultones	No
Epoxides and aziridines	No
Aliphatic halogens	No
Alkyl nitrite	No
Simple aldehyde	No

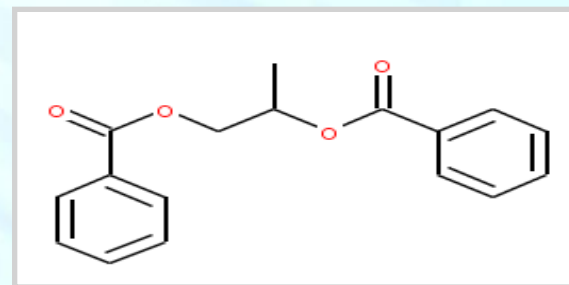
Toxtree internals

Series of questions are applied on query compound and also on reaction products

Q1.Normal constituent of the body No
Q2.Contains functional groups associated with enhanced toxicity No
Q3.Contains elements other than C,H,O,N, divalent S No
Q5.Simply branched aliphatic hydrocarbon or a common carbohydrate No
Q6.Benzene derivative with certain substituents No
Q7.Heterocyclic No
Q16.Common terpene No
Q17.Readily hydrolysed to a common terpene No
Q19.Open chain No
Q23.Aromatic Yes
Q27.Rings with substituents Yes
Q28.More than one aromatic ring Yes
Q29.Readily hydrolysed Yes

Q19.Open chain Yes
Q20.Aliphatic with some functional groups Yes
Q21.3 or more different functional groups No
Q18.One of the list (see explanation) No **Low (Class I)**

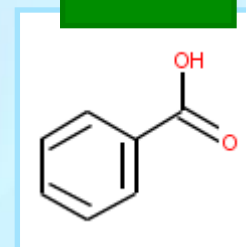
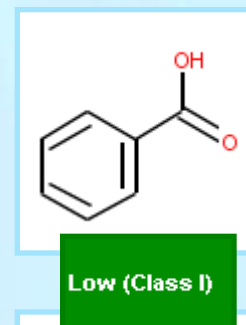
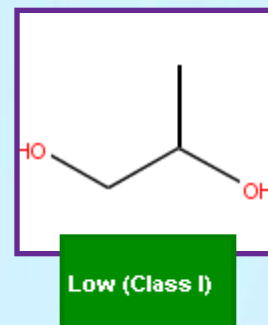
Q30.Aromatic Ring with complex substituents No
Q18.One of the list (see explanation) No **Low (Class I)**



Query compound

Hydrolysis

products



<http://ambit.sourceforge.net>

AMBIT : Building blocks for a (Q)SAR decision support system

AMBIT 2.0

Download

Project Documentation

Modules

ambit2-base
ambit2-core
ambit2-hashcode
ambit2-smarts
ambit2-descriptors
ambit2-ds
ambit2-smi2d
ambit2-mopac
ambit2-ai
ambit2-dhui
ambit2-workflow
ambit2-plugin: Database search and Analogue identification
ambit2-plugin: Category building
ambit2-plugin: Database administration
ambit2-namestructure
ambit2-plugin: REACH PBT assessment
ambit2-taglib
Pulchem utilities
ambit2-REST web services
ambit2-model

sourceforge

Download

AMBIT XT

Video guides: Read across using SMARTS substructure and similarity search PBT assessment

AmbitXT v 2.2.0

- November 2009 : Download AmbitXT 2.2.0 release
 - Windows setup with an embedded MySQL database
 - Integrated Toxtree (Toxtree modules available from the Descriptor calculation interface);
 - included chemical structure data, gathered from several independent sources;
 - Improved performance;
 - Updated Toxtree modules
 - Technical Guide
 - User Manual
 - Manual for PBT assessment module
- The included database encompasses substances from the following sources:
 - ECHA Preregistration list
 - Bioconcentration factor (BCF) Gold Standard Database
 - ECETOC Aquatic Toxicity (EAT) Database Supplement to ECETOC, 2003. Aquatic Hazard Assessment II. Technical Report No. 91. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium
 - Compilation of historical local lymph node assay data for the evaluation of skin sensitization alternatives. Gerberick GF, Ryan CA, Kern PS, Schlatter H,

Search substances in REPOSE - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://nina.acad.bg/item/search.jsp

Results appearance Structures per page 10

Identification

CAS Registry number Search

Structure

Structure diagram Search

Similarity

Structure diagram Search

Tanimoto

Ambit Database Tools 1.30

File Database Molecule Search MySQL Help

Substructure

Enter SMILES

Molecule browser

78 molecules

#	Structure diagram	CasRN	Names
1		108-95-2	PHENOL
2		123-31-9	Hydroquinone
3		108-46-3	Resorcinol

Identifiers Descriptors Structure Experimental data Metabolites

Chemical name
Structure search
Molecular formula
Search AQUIRE
Retrieve dataset
Batch search
CAS RN
Aliases
SMILES

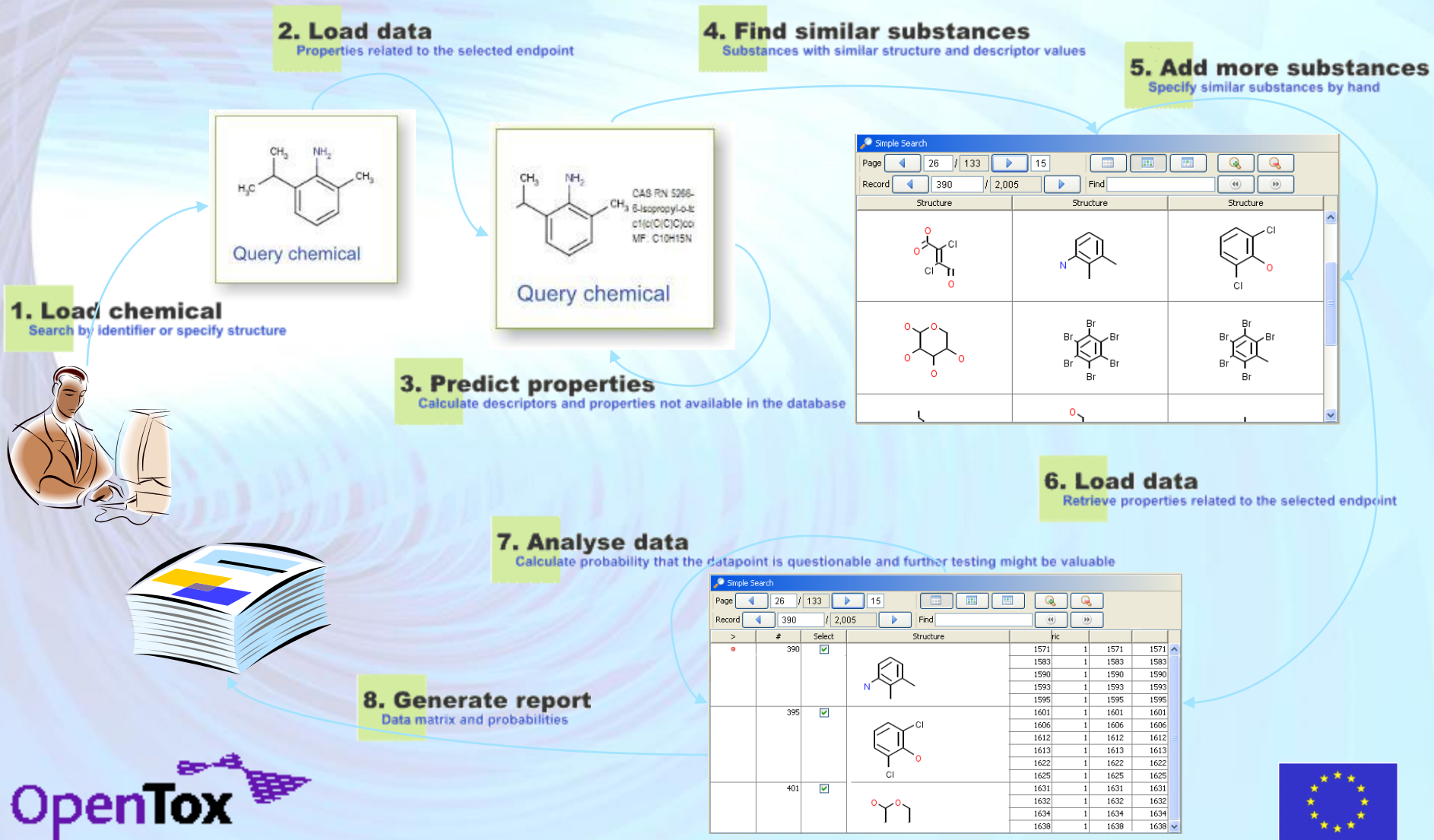
Search options

Dataset: ALL
Options

Build 3D completed.

Processed 1 records in 27 seconds (27 s per record)

A workflow in AMBIT XT



AMBIT Extensions

<http://qsardb.jrc.it>

- EC JRC Institute for Health and Consumer Protection (former European Chemicals Bureau) - a reference site for retrieving robust summaries of (Q)SAR models in QSAR Model Reporting Format (QMRF)
- AMBIT 2.0
- RESTful Web services
- Custom extensions for third parties

OpenTox

The top screenshot shows the 'QMRF documents' page in Mozilla Firefox. The page title is '(Q)SAR Model Reporting Format Inventory'. It features a navigation bar with 'Home', 'Search documents', 'Search structures', 'My documents', 'New document', and 'My profile'. A table lists documents with columns: #, QMRF#, Title, Version, Author, Last updated, Status, View, Download, and Actions. The table shows one document with QMRF# 1, Title 'SSSSSSS', Version 1, Author 'nina', Last updated '2008-12-24 09:51', and Status 'draft'. The bottom screenshot shows the 'QMRF Database: Upload QMRF document attachments' page. It features a table for uploading attachments with columns: #, File, File format, Description, Type, and Updated. The table shows one attachment with File 'ECETOC_skin_scores_structures_validated_IJCHP.sdf', File format 'sdf', Description 'test', Type 'data_validation', and Updated '2009-01-27 16:54:57.0'. Below this table is a section for 'Attachment' with columns: #, Properties, Import as (type), and Import as (name). The table shows 14 attachments, all with Properties 'Erythema (rabbit 1)-1d', 'Erythema (rabbit 1)-1h', 'Erythema (rabbit 1)-2d', 'Erythema (rabbit 1)-3d', 'Erythema (rabbit 1)-7d', 'Erythema (rabbit 2)-1d', 'Erythema (rabbit 2)-1h', 'Erythema (rabbit 2)-2d', 'Erythema (rabbit 2)-3d', and 'Erythema (rabbit 2)-7d', and Import as (type) 'ignore'.

OpenTox summary

- HEALTH-2007-1.3.3 Promotion, development, acceptance and implementation of QSARs (quantitative structure-activity relationship) for toxicology
- 11 partners; <http://opentox.org>
- The overall objective of the proposed project is to develop a framework, that provides a unified access to toxicity data, predictive models, procedures supporting validation and additional information that helps with the interpretation of predicted results.
- The OpenTox framework is accessible at three levels:
 - A simple and intuitive interface for toxicological experts, that provides unified access to predictions, toxicological data, models and supporting information
 - An expert interface for the streamlined development and validation of new models
 - An application programming interface (API) for the development, integration and validation of new algorithms

OpenTox Is A Framework

Framework

- Toxicity data
- QSAR models
- Validation support
- Interpretation aids

Unified Access

- Toxicologists
- QSAR Modelers
- API for new QSAR algorithm development & integration

Open Source

- To optimise impact
- To allow inspection / review
- To attract external contributors

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	well defined data and access to data, unambiguous vocabulary
2	Unambiguous Algorithm	providing unified access to documented models and algorithms as well as to the source code of their implementation
3	Defined Applicability Domain	integrating tools for the determination of applicability domains and considering these during the validation of (Q)SAR models
4	Goodness-of-fit, robustness and predictivity	providing scientifically sound validation routines for the determination of these measures
5	Mechanistic interpretation (if possible)	prediction of toxicological mechanisms; inclusion of biological data; extensibility

Consequences for Requirements on OpenTox Framework

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>

Challenges for integration platform in predictive toxicology: The Data

- Flat text files (legacy formats)
- Data type is implicit and part of the data (no metadata)
 - > <TargetSites_Mouse_Male>
no positive results
- Multiline records
 - > <Endpoint>
TD50
Tumor Target Sites
- Records duplication
 - > <Species>
 - rat
 - 489 occurrences in
CPDBAS_v5c_1547_29Apr2008.sdf
- Some records have internal structure (subrecords)
 - > <Species>
 - rat, mouse
- Records order matters

Good

- Simplicity
- Manual inspection
- Sequential processing

Bad

- Data standardization
- Data summarization
- Data duplication
- Metadata

Ugly

- Random access processing
- Huge volumes processing
einecs_structures_V13Apr07.sdf 135MB
- Records modification
- Errors !

Challenges for integration platform in predictive toxicology: Chemical structures and Models

- Journal publications
- Commercial software
- Open source software
- No common means of sharing models
- Recent attempt to formalize description of models (QMRf by JRC)

Good

- Many models available
- >20 000 QSARs
- >5000 descriptors

Bad

- Hard or impossible to reproduce published models
- Ambiguity in chemical structures
- Errors in chemical structures

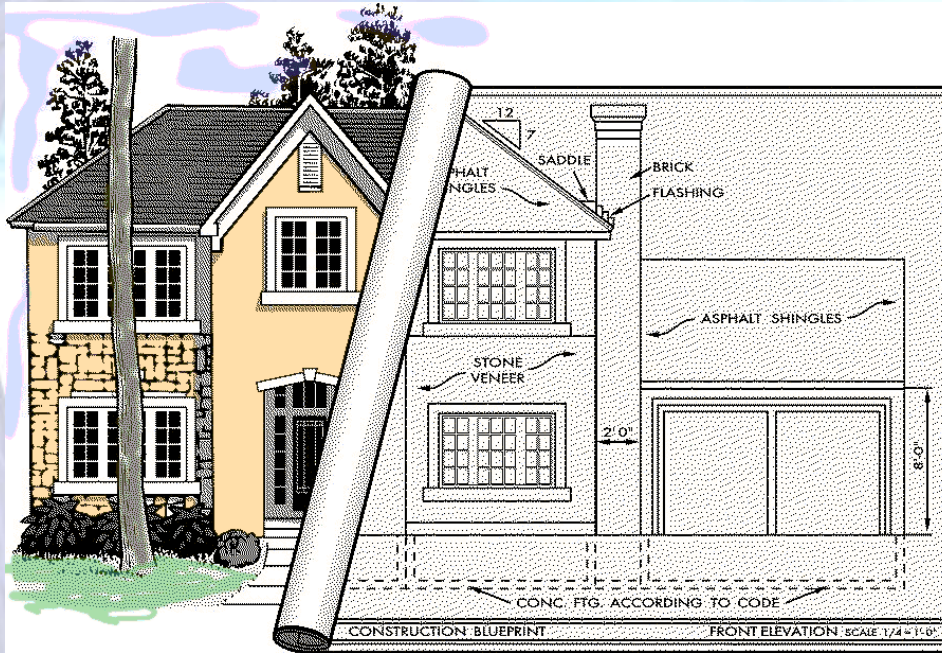
Ugly

- Information loss and efforts duplication
- Same chemical structures are submitted to (almost) the same processing

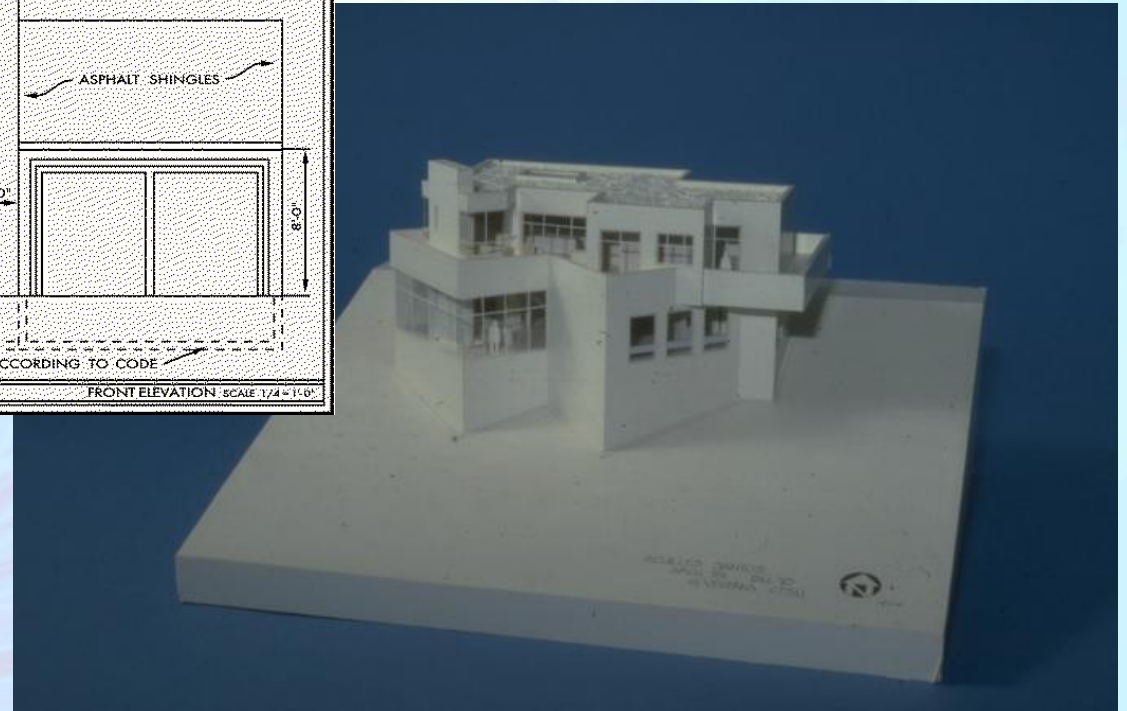
Software Architecture



Software Architecture



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



OpenTox

- Distributed Web services for predictive toxicology
- REST technology
 - Every object has an unique URI
 - URIs are dereferensable
 - Multiple representation of an object is encouraged (e.g. RDF, but also others)
 - Fixed operations - GET, PUT, POST, DELETE
- Every object has RDF representation
 - Compounds
 - Datasets
 - Compound properties
 - Prediction algorithms
 - Models
 - Validation statistics
 - Reports
- Ontologies: Opentox.owl, Blue Obelisk algorithm ontology, OpenTox algorithm types ontology, OpenTox endpoints ontology, based on ECHA endpoints classification; specific endpoints ontologies, developed by ISS &

Representational State Transfer (REST): What and Why?

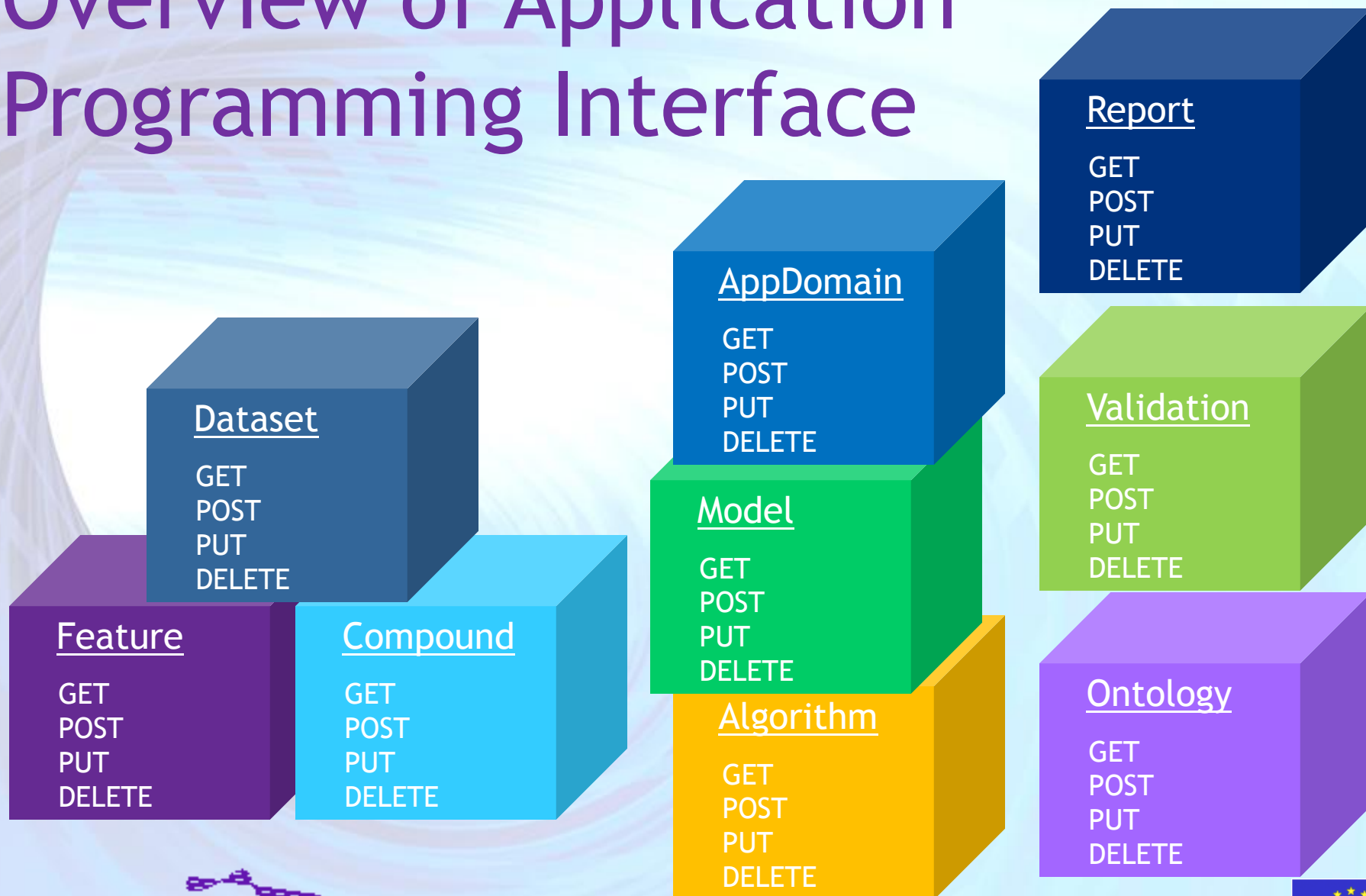
What?

- 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

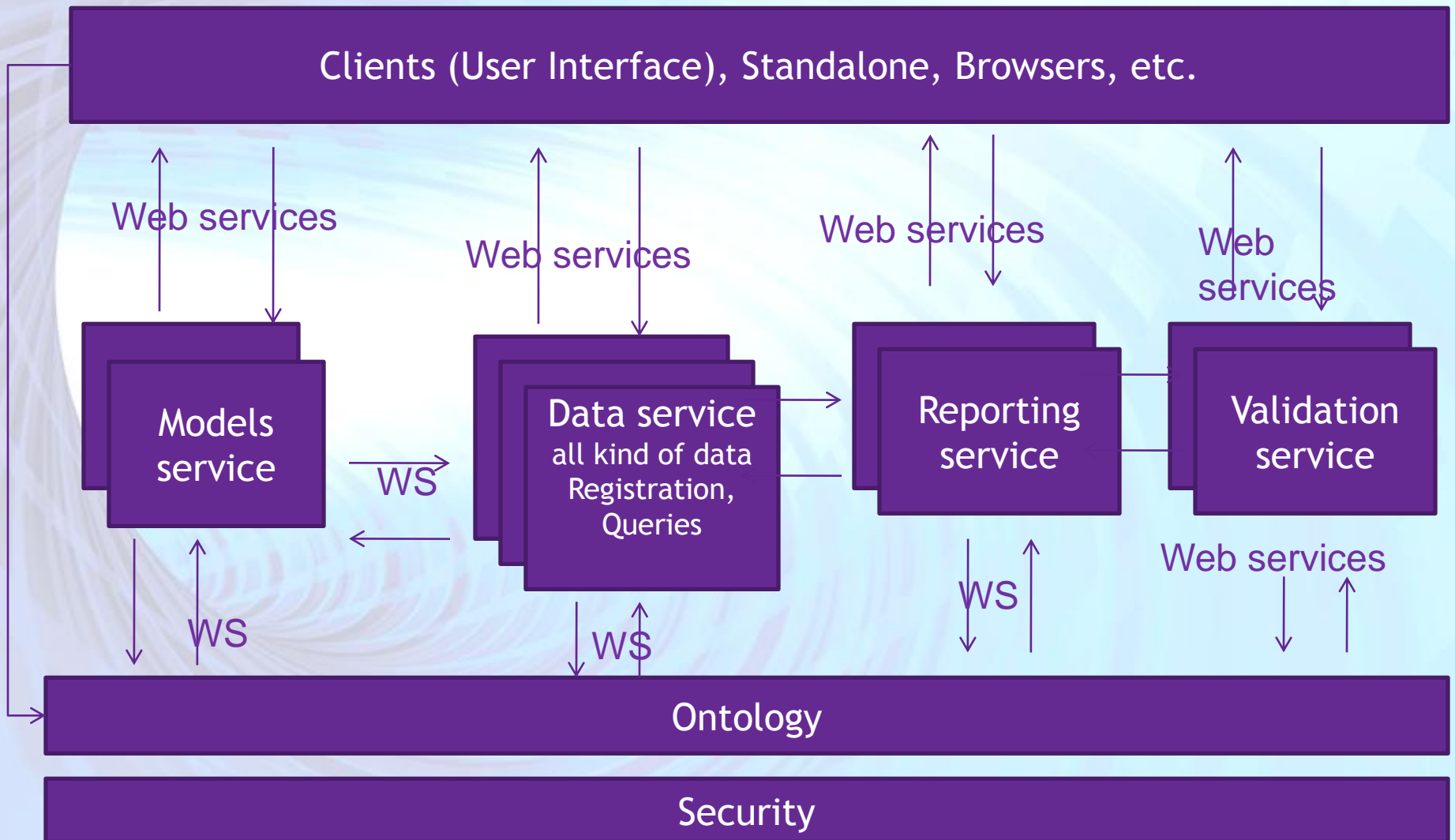
Overview of Application Programming Interface



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

Information exchange based on standardized ontologies and REST web services. All the information from different services might be used separately or via an ontology repository with reasoning capabilities



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 proposals for tox ontologies out there
- Our strategy: use existing work and standards wherever possible
- However, new ontology, e.g., for algorithms

CLASS BROWSER

For Project:  AlgorithmTypes

Class Hierarchy    

- owl:Thing (14)
 - ota:AlgorithmType
 - ota:DescriptorCalculation
 - ota:PatternMining (2)
 - ota:PharmacophoreGeneration
 - ota:PhysicoChemical
 - ota:QuantumChemical
 - ota:SimilarityDistance
 - ota:Topological
 - ota:MSDMTox
 - ota:Clustering
 - ota:Learning
 - ota:Classification (4)**
 - ota:EagerLearning (4)
 - ota:LazyLearning (4)
 - ota:Regression (4)
 - ota:MultipleTargets (4)
 - ota:Rules
 - ota:SingleTarget (4)
 - ota:Preprocessing
 - ota:DataCleanup
 - ota:Discretization (2)
 - ota:FeatureSelection (2)
 - ota:Normalization
 - ota:SemiSupervised
 - ota:Supervised (11)
 - ota:Unsupervised (3)
 - ota:Utility
 - ota:Generation3D
 - ota:SimilarityDistanceCalculation
 - ota:Visualisation

OpenTox object ontology

<http://www.opentox.org/api/1.1/opentox.owl>

opentox Protégé 3.4.1 (file:\E:\src-exp\ontology\src\main\resources\org\opentox\owl\opentox.pprj, OWL / RDF Files)

File Edit Project OWL Reasoning Code Tools Window Collaboration Help

Metadata() OWLClasses Properties Individuals Forms

SUBCLASS EXPLORER
For Project: opentox

Asserted Hierarchy

- owl:Thing
 - ot:OpentoxResource
 - ot:Algorithm
 - ot:NumericAlgorithm
 - ot:Compound
 - ot:Conformer
 - ot:Crossvalidation
 - ot:DataEntry
 - ot:DataEntryNumeric
 - ot:Dataset
 - ot:EmptyDataset
 - ot:NumericDataset
 - ot:DataType
 - ot:Nominal
 - ot:Numeric
 - ot:String
 - ot:Feature
 - ot:NominalFeature
 - ot:NumericFeature
 - ot:StringFeature
 - ot:FeatureValuePair
 - ot:Model**
 - ot:Parameter
 - ot:Report
 - ot:Task
 - ot:Validation
 - ot:ValidationInfo
 - ot:ClassificationStatistics

CLASS EDITOR for ot:Model (instance of owl:Class)
For Class: <http://www.opentox.org/api/1.1#Model> ☐ Inferred View

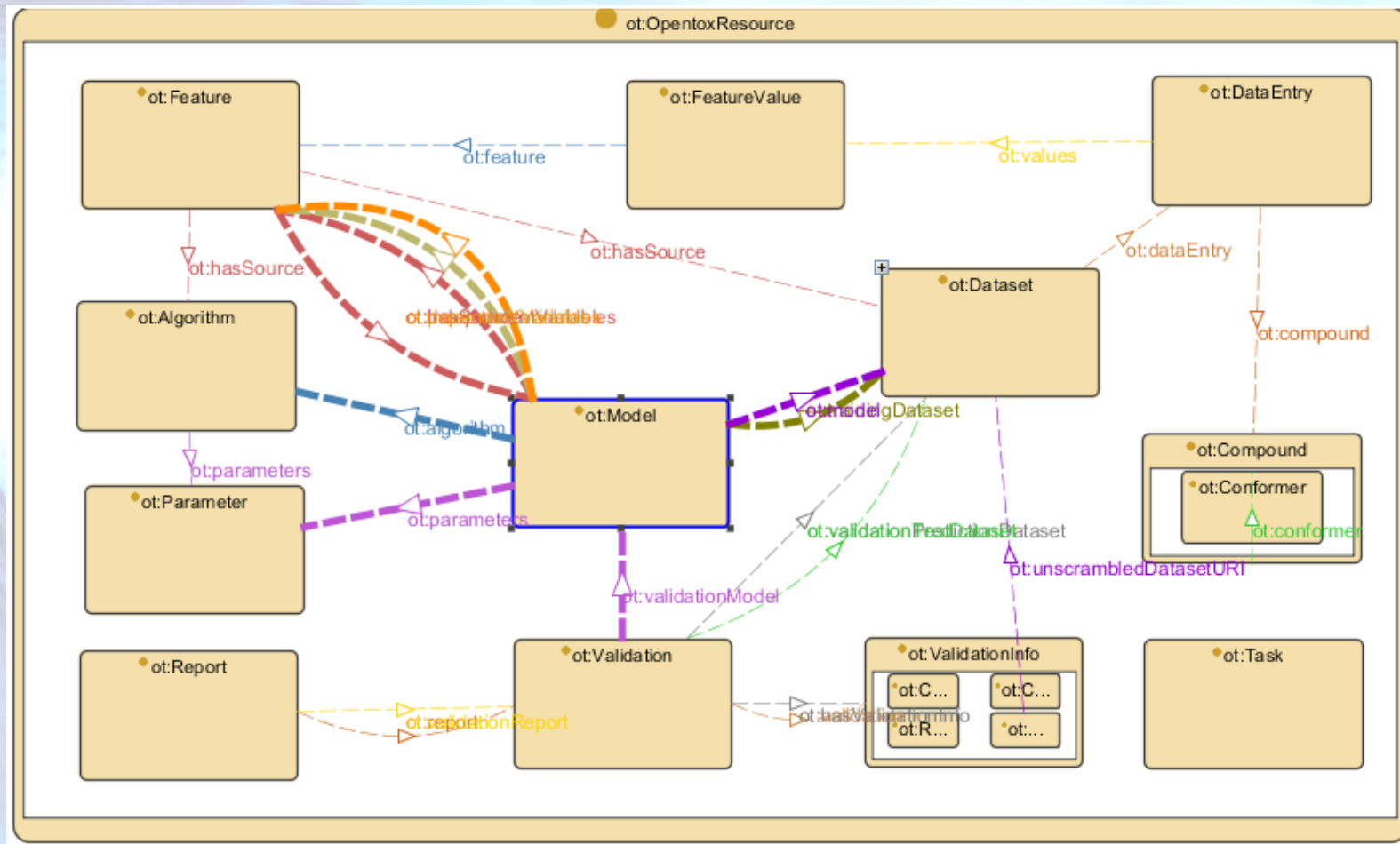
Property	Value
rdfs:comment	
dc:creator	The model creator (perhaps a link to User resource)
dc:date	The date of model creation
dc:format	The native format of the model content (e.g. PMML, Weka model, etc.)

Subclasses

- ot:OpentoxResource
- ot:Parameter

OpenTox object ontology

<http://www.opentox.org/api/1.1/opentox.owl>



Toxicity endpoints ontology, based on ECHA classification

<http://www.opentox.org/api/1.1/echa-endpoints.owl>

The screenshot shows the Protégé 3.4.1 interface with the 'echa-endpoints' ontology loaded. The left pane, 'Subclass Explorer', shows the hierarchy of classes under 'otee:Endpoints'. The right pane, 'Class Editor', shows the properties of the selected class 'otee:HumanHealthEffects'.

Subclass Explorer (Left Pane):

- otee:Endpoints
 - otee:EcotoxicEffects
 - otee:Acute_toxicity_to_fish_letality
 - otee:Long-term_toxicity_to_Daphnia_letality_inhibition_of_reproduction
 - otee:Long-term_toxicity_to_fish_egg_sac_fry_growth_inhibition_of_juvenile_fish_early_life_stage
 - otee:Microbial_inhibition_activated_sludge_respiration_inhibition_inhibition_of_nitrification_other
 - otee:Short-term_toxicity_to_algae_inhibition_of_the_exponential_growth_rate
 - otee:Short-term_toxicity_to_Daphnia_immobilisation
 - otee:Toxicity_to_birds
 - otee:Toxicity_to_earthworms_survival_growth_reproduction
 - otee:Toxicity_to_plants_leaves_seed_germination_root_elongation
 - otee:Toxicity_to_sediment_organisms_survival_growth_reproduction
 - otee:Toxicity_to_soil_invertebrates_survival_growth_reproduction
 - otee:Toxicity_to_soil_microorganisms_inhibition_of_C-mineralisation_inhibition_of_N-mineralisation
- otee:EnvironmentalFateParameters
- otee:HumanHealthEffects
 - otee:Acute_photolrritation
 - otee:AcuteDermalToxicity
 - otee:AcuteInhalationToxicity
 - otee:AcuteOralToxicity
 - otee:Carcinogenicity
 - otee:EndocrineActivity
 - otee:Other_e.g._inhibition_of_specific_enzymes_involved_in_hormone_synthesis_or_regulation
 - otee:Receptor-binding
 - otee:Receptor_binding_and_gene_expression
 - otee:Eye_irritation_corrosion
 - otee:In_vitro_reproductive_toxicity_e.g._embryotoxic_effects_in_cell_culture_such_as_embryo_survival

Class Editor (Right Pane):

For Class: `.org/echaEndpoints.owl#HumanHealthEffects`

Property	Value
rdfs:comment	
identifier	4
title	Human health effects

Ontology service

- RDF triple storage
- REST interface for registration of OpenTox objects
- SPARQL query

Search OpenTox RDF - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://apps.ideaconsult.net:8080/ontology/query/Endpoints

Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options

Search OpenTox RDF

Features Algorithms Models Endpoints

Import RDF data into Ontology service

URL

SUBMIT

Ontology service 17969 triples

SPARQL

```
PREFIX ot:<http://www.opentox.org/api/1.1#>
PREFIX otee:<http://www.opentox.org/echaEndpoints.owl#>
select ?Endpoints ?title ?id
where {
  ?Endpoints rdfs:subClassOf otee:Endpoints.
  OPTIONAL (?Endpoints dc:title ?title).
  OPTIONAL (?Endpoints dc:identifier ?id).
}
```

Submit Query

Results [found in 1 ms]

Endpoints	title	id
http://www.opentox.org/echaEndpoints.owl#PhysicoChemicalEffects	Physicochemical effects ""http://www.w3.org/2001/XMLSchema#string	"1""http://www.w3.org/2001/XMLSchema#string
http://www.opentox.org/echaEndpoints.owl#Toxicokinetics	Toxicokinetics ""http://www.w3.org/2001/XMLSchema#string	"5""http://www.w3.org/2001/XMLSchema#string
http://www.opentox.org/echaEndpoints.owl#EcotoxicEffects	Ecotoxic effects""http://www.w3.org/2001/XMLSchema#string	"3""http://www.w3.org/2001/XMLSchema#string
http://www.opentox.org/echaEndpoints.owl#EnvironmentalFateParameters	Environmental fate parameters ""http://www.w3.org/2001/XMLSchema#string	"2""http://www.w3.org/2001/XMLSchema#string
http://www.opentox.org/echaEndpoints.owl#HumanHealthEffects	Human health effects""http://www.w3.org/2001/XMLSchema#string	"4""http://www.w3.org/2001/XMLSchema#string

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

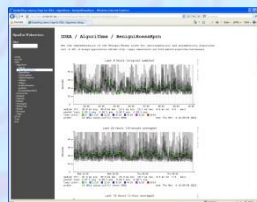
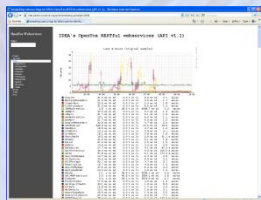
OntologyWS: IDEA

OpenTox database supported via Dataset service

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

Implementation: key facts (1)

- Seven independent open source implementations of different OpenTox API-v1.1 subsets have been developed in three different programming languages (Java - 4, Ruby - 2, Delphi - 1);
- Continuous availability and performance monitoring of a selected subset of OpenTox webservices (including integration aspects), running and gathering detailed statistics since December 2009;
- **Next major milestone - design and implementation of authentication and authorisation (AA) support:**



Use case 1): 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](#).

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

name for your endpoint:

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

Browse...

Create model Cancel

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

Use case 2: ToxPredict

<http://toxpredict.org>

- ToxPredict estimates the chemical hazard of chemical structures. It relies on [OpenTox API-v1.1](#) compliant RESTful webservice.
- Users can either search the OpenTox prototype database, which includes currently quality labelled data for **~150,000 chemicals**, grouped in more than a dozen datasets, or upload their own chemical structure data. ToxPredict provides access to **14 ready to use models**, addressing **14 different endpoints** (and growing!);

The screenshot displays the ToxPredict web application in a Mozilla Firefox browser. The page title is "ToxPredict" and it is identified as an "OpenTox demo application". A navigation bar at the top right includes links for "Welcome", "guest", "Admin", and "Help". A central workflow bar shows five steps: 1. Select structure(s), 2. Verify structure(s), 3. Select model(s), 4. Run prediction(s), and 5. Display result(s). Below this, a message states: "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, SM, PDF, CSV, ARFF, RDF/XML or RDF/NS." A "Hide" button is present. The main content area shows "Page 0 Records per page 10" and "Structure(s) & Model predictions". It displays a chemical structure of 2-(4-methylphenyl)propanal (CASRN 99-72-9) with its SMILES string: CC(C=O)c1ccc(C)cc1. Below the structure, a table lists 14 endpoints and their predicted values.

Endpoint	Dataset/Model	Name	Value	Predicted/Experimental
http://www.opentox.org/api/v1.1/LipinskiFailures	http://apps.ideaconsult.net:8080/ambit2algorithm/org.openscience.cdk.qsar.descriptors.molecular.RuleOfFiveDescriptor	LipinskiFailures	0.0	Predicted
Carcinogenicity	http://apps.ideaconsult.net:8080/ambit2model8	For a better assessment a QSAR calculation could be applied.	NO	Predicted
Carcinogenicity	http://apps.ideaconsult.net:8080/ambit2model8	Negative for genotoxic carcinogenicity	NO	Predicted
Carcinogenicity	http://apps.ideaconsult.net:8080/ambit2model8	Negative for nongenotoxic carcinogenicity	YES	Predicted

ToxPredict: Step 1 (Select structure(s))



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

OT Dataset API *HTTP GET*

OT Dataset Service

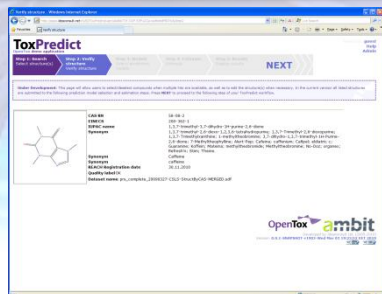
ToxPredict Web Application

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 (Select model(s))

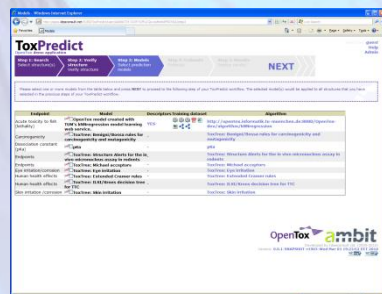


ToxPredict
Web
Application

HTTP GET SPARQL query

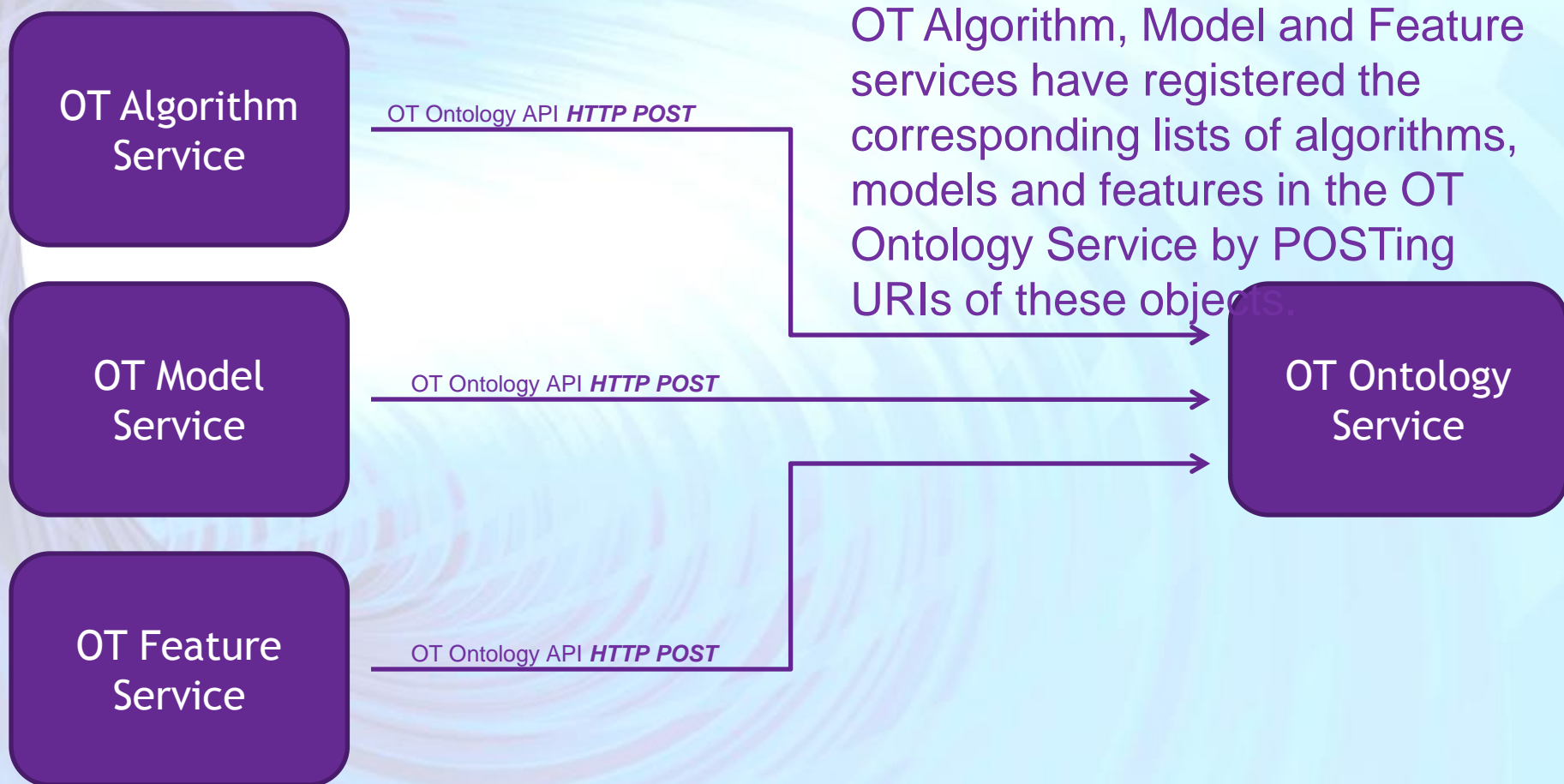
OT Ontology
Service

application/sparql-
results+xml

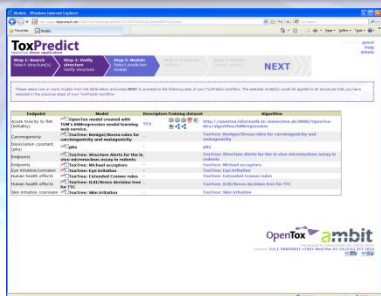


Here is the list of model URIs and
related endpoints and algorithms
in SPARQL format.

ToxPredict: Step 3 (behind the scenes)



ToxPredict: Step 4 (Estimate)



Run the selected models.

OT Model API **HTTP POST** with

ToxPredict
Web
Application

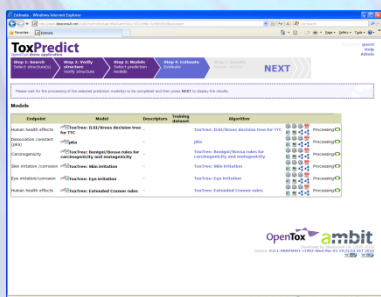
OT Model Service

HTTP code 202 "Accepted"

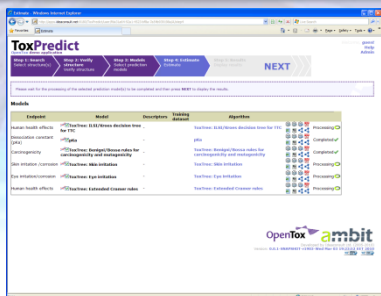
The calculation will take a while; here is a task URI, which can be queried for processing status.

Create a new task.

OT Task Service



ToxPredict: Step 4 (Estimate)



ToxPredict
Web
Application

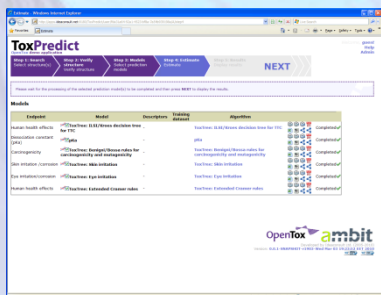
Is the task completed?

OT Task API **HTTP GET** on task URL

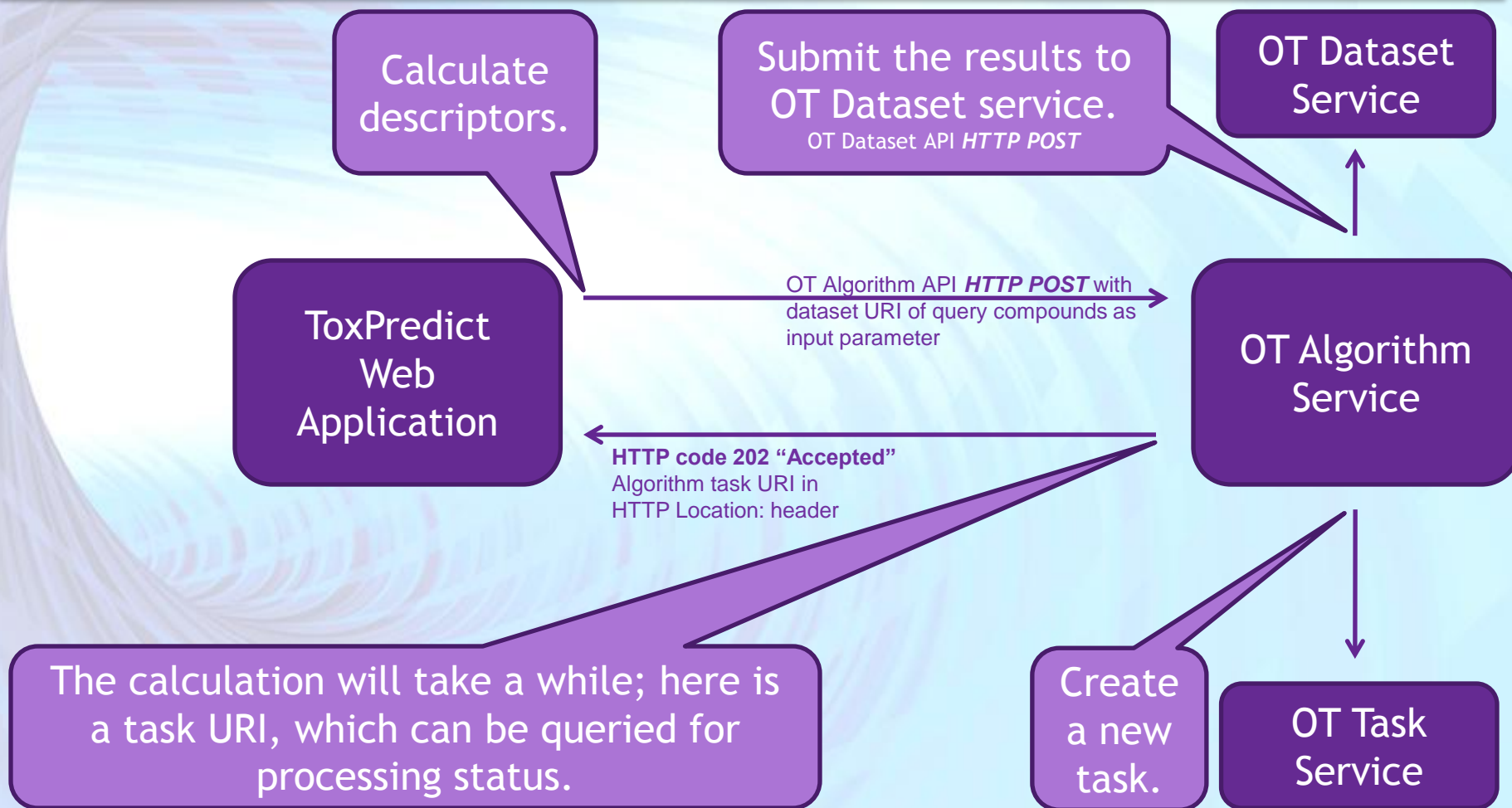
OT Task Service

← HTTP code 200 “OK”

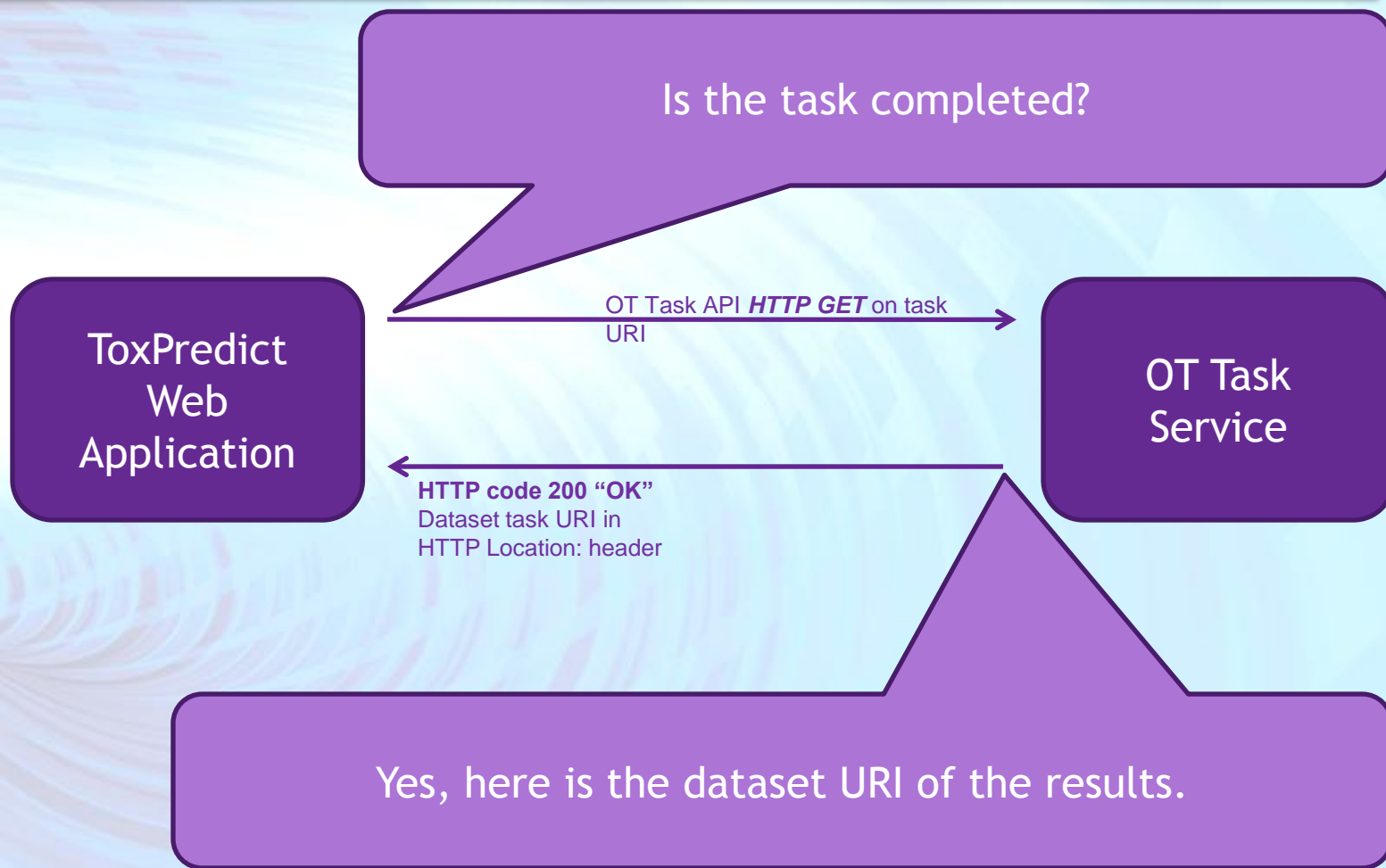
Yes, here is the dataset URI of the results.



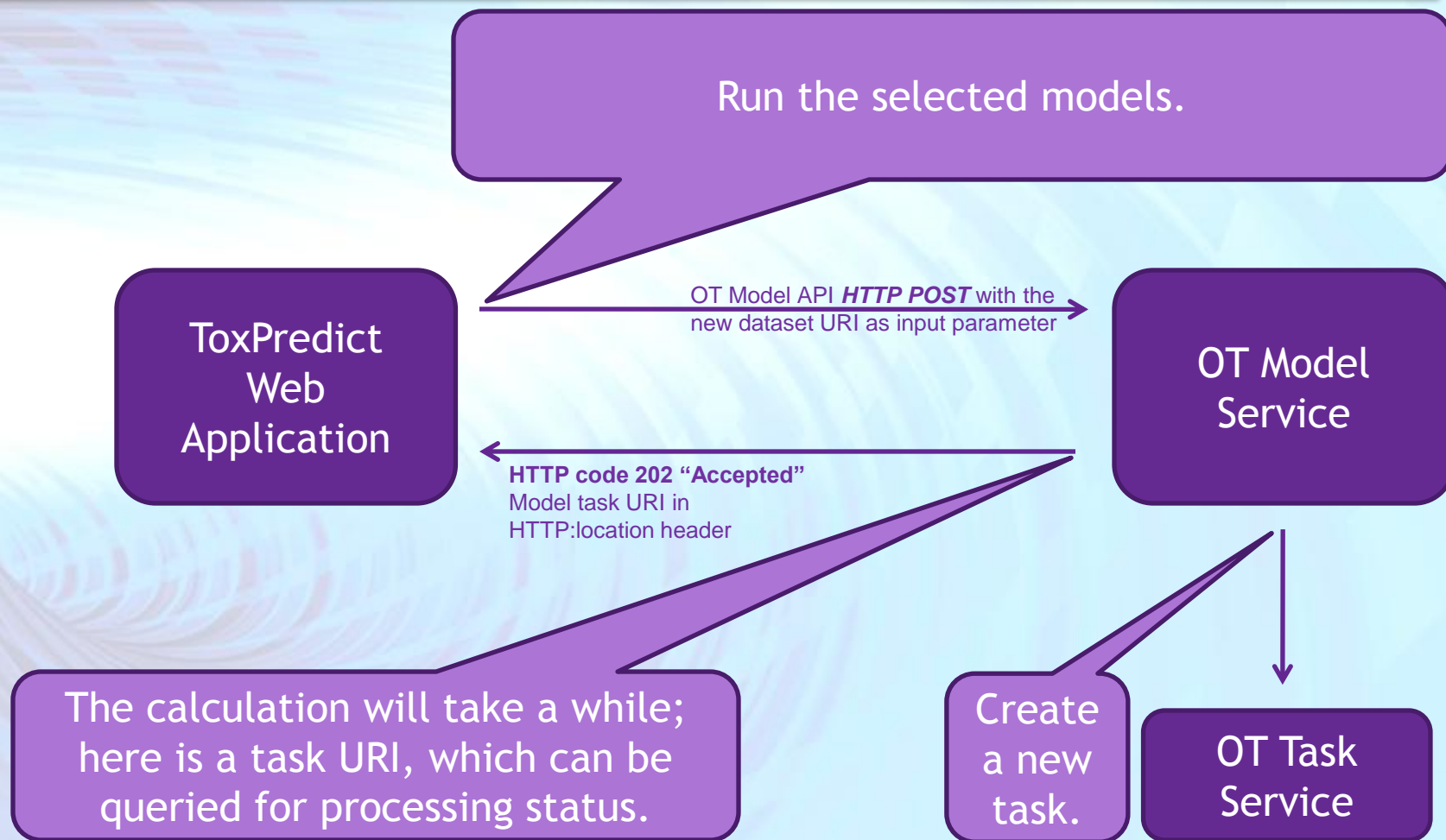
ToxPredict: Step 4 (behind the scenes)



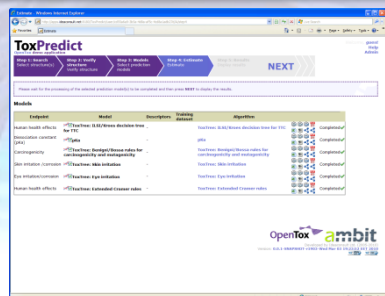
ToxPredict: Step 4 (behind the scenes)



ToxPredict: Step 4 (behind the scenes)



ToxPredict: Step 5 (Display results)



Retrieve calculation results from the final dataset URI, obtained in Step 4 (Estimation).

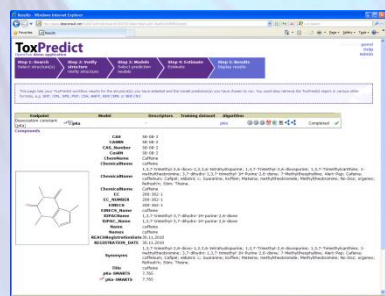


ToxPredict
Web
Application

OT Dataset API *HTTP GET*

OT Dataset
Service

application/rdf+xml



Here is the dataset content in RDF format, according to OpenTox.owl and containing estimation results, as well as compound identifiers and experimental data.

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: e.g.,
 - Infrastructure and API for providing restricted access to data and models
 - Improvements in ToxCreate and ToxPredict use cases
 - Read-across (technical infrastructure available),
 - ToxCast analysis
 - Pathway prediction service: developing a service for predicting transformation products

Do you develop models and algorithms for prediction of biological effects?

It can be made available as OpenTox web service!

Thank You For Your Attention!

OpenTox events:

- 1 day OpenTox workshop at AXLR8 meeting (Potsdam, Germany, May 30 2010)
 - eChemInfo hands-on training Cambridge, UK, Aug 2010
- Oral & Poster Presentation @ ACS National Meeting , section *Semantic Web in Chemistry*, Boston, MA, USA, Aug 2010
- 1 day OpenTox workshop at EuroQSAR 2010 (Rhodos, Greece, Sep 2010)