OpenTox

Dr. Nina Jeliazkova
Ideaconsult Ltd. Sofia, Bulgaria
Why integration framework for predictive toxicology?
Why integration framework for predictive toxicology?

- **Challenges:**
  - Chemical structures
    - Might be ambiguous
    - Might be error prone or time consuming to reproduce from publications
  - Data
    - **Multiple formats,**
    - **Implicit semantics,** often buried in human readable documentation only
  - Models
    - **Tens of thousands** available, in software or in publications
    - Multiple software solutions, mostly incompatible
    - Predictions **reproducibility** is time consuming and often hard to achieve
    - Automatic **comparison of prediction results** difficult
## OpenTox Framework approach

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

- Distributed Web Services for predictive toxicology
- Several types of Web Services (using REST Web service technology)
- Service types corresponds to following building blocks:
  - Chemical compound;
  - Dataset of chemical compounds and their properties (calculated or measured)
  - Algorithm (descriptor calculation, regression, classification, structural alerts, QC, etc.)
  - Predictive model
  - Report, validation, applicability domain, etc.

- Every object (compound, dataset, algorithm, model, etc.) has an unique web address (e.g. http://myhost.com/model/bestpredictivemodel)
- These objects can be created, read, deleted, and updated
- Every object has RDF (W3C Resource Description Framework) representation, defined in OpenTox ontology (opentox.owl)
OpenTox components
Overview of OpenTox API
(Application Programming Interface)

• The way applications talk to each other
• The way developers talk to applications
### OpenTox datasets: Unified access to data

Everything described by W3C RDF (Resource Description framework)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://myhost.com/compound/413">http://myhost.com/compound/413</a></td>
<td>N,N-dimethyl-4-aminoazobenzene</td>
<td>CN(C1=CC=C(C=C1)N=N/C2=CC=CC=C2)C</td>
<td>3</td>
<td>3.31</td>
<td>225.3</td>
<td>YES</td>
<td>3.123</td>
<td></td>
</tr>
<tr>
<td><a href="http://myhost.com/compound/44497">http://myhost.com/compound/44497</a></td>
<td>4-acetamidofluorene</td>
<td>O=C(Nc3c2c1ccc(cc1)Cc2ccc3)C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

---

```xml
http://myhost.com/feature/21573
http://myhost.com/feature/21858
http://myhost.com/feature/22114

a ot:Feature, ot:NumericFeature ;
dc:creator "http://www.blueobelisk.org/ontologies/chemoinformatics-algorithms/#xlogP" ;
dc:title "XLogP" ;
ot:hasSource <http://myhost.com/algorithm/org.openscience.cdk.qsar.descriptors.molecular.XLogPDescriptor> ;
= otee:Octanol-water_partition_coefficient_Kow .
```
Uniform access to the data

- Ontologies are critical to unambiguously describe data
  - Opentox.owl;
  - Blue Obelisk algorithm ontology (cheminformatics algorithms);
  - OpenTox algorithm types ontology;
  - OpenTox endpoints ontology, based on ECHA endpoints classification;
  - Specific endpoints ontologies, developed by OpenTox partners with toxicology expertise;
  - Existing biomedical ontologies.
- Datasets can be easily merged, compared, and calculations reproduced, regardless of their physical place.
- OWL sameAs construct can be used to denote two links point to the same object.
- Authentication and authorization to protect sensitive data
Uniform access to the data

- Datasets can be easily merged, compared, and calculations reproduced, regardless of their physical place.
- The dataset service offers property, compound, substructure and similarity searches via uniform OpenTox Application Programming Interface.
Example: mutagenicity dataset

- [http://apps.ideaconsult.net:8080/ambit2/dataset/2344](http://apps.ideaconsult.net:8080/ambit2/dataset/2344) (the dataset)
Example: mutagenicity dataset

```turtle
@prefix ot: <http://www.opentox.org/api/1.1#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix : <http://apps.ideaconsult.net:8080/ambit2/> .
@prefix ota: <http://www.opentox.org/algorithmTypes.owl#> .
@prefix otee: <http://www.opentox.org/echaEndpoints.owl#> .
...
@prefix af: <http://apps.ideaconsult.net:8080/ambit2/feature/> .

af:28958
  a ot:Feature , ot:NumericFeature .
  dc:creator "194.141.0.136" ;
  dc:title "Activity" ;
  ot:hasSource "tox_benchmark_N6512.sdf" ;
  ot:units "" ;
  = otee:Mutagenicity .

ot:hasSource
  a owl:ObjectProperty .

ot:units
  a owl:DatatypeProperty .

ot:Feature
  a owl:Class .

ot:NumericFeature
  a owl:Class ;
  rdfs:subClassOf ot:Feature .
```
Query: Is there other mutagenicity data available?

Merge mutagenicity data

Uniform access to calculations

Read data from a web address - process - write to a web address

HTTP requests:
- GET
- POST
- PUT
- DELETE

Dataset
- GET
- POST
- PUT
- DELETE

Compound
- GET
- POST
- PUT
- DELETE

Model
- GET
- POST
- PUT
- DELETE

Predicted results

http://myhost.com/dataset/newcompounds
http://myhost.com/model/predictivemodel1
http://myhost.com/dataset/predictedresults1
List available models

- http://apps.ideaconsult.net:8080/ambit2/model
Apply ToxTree module for carcinogenicity and mutagenicity prediction

- Clicking ‘Predict’ button initiates calculations
- It actually sends HTTP POST command to the model URL, with parameter the dataset URL

- The benchmark dataset web address
Long running calculations - web address again

- http://apps.ideaconsult.net:8080/ambit2/task/e75057a0-b005-4c74-949f-0be9304a3c42/model
And the results (a dataset web address)


Structural Alert for genotoxic carcinogenicity
Building blocks for model creation

- A Dataset
  - Datasets can be uploaded (SDF, CSV, SMI, MOL)
    - via HTTP POST to the dataset service
    - Web browser file upload
  - Existing datasets (uploaded by others) can be used
  - Dataset Web address assigned

- Descriptor calculation
- Learning algorithm
- Perform predictions
- Make the model available
The dataset (Caco-2 dataset from a JCIM publication)

- http://apps.ideaconsult.net:8080/ambit2/dataset/54
- http://apps.ideaconsult.net:8080/ambit2/dataset/54/metadata
- The endpoint (caco2 column)
  http://apps.ideaconsult.net:8080/ambit2/feature/22200
Uniform approach to data processing (e.g. Descriptors calculation)

Read data from a web address - process - write to a web address

Dataset
- GET
- POST
- PUT
- DELETE

Feature
- GET
- POST
- PUT
- DELETE

Compound
- GET
- POST
- PUT
- DELETE

Algorithm
- GET
- POST
- PUT
- DELETE

Dataset
- GET
- POST
- PUT
- DELETE

Feature
- GET
- POST
- PUT
- DELETE

Compound
- GET
- POST
- PUT
- DELETE

http://myhost.com/dataset/trainingset1

http://myhost.com/dataset/results

http://myhost.com/algorithm/neuralnetwork

OpenTox
March 17, 2011
The algorithms (including descriptors calculation)

- http://apps.ideaconsult.net:8080/ambit2/algorith
Descriptors calculation


- Clicking ‘Predict’ button initiates calculations
- It actually sends HTTP POST command to the algorithm URL, with parameter the dataset URL

- The demo dataset web address
- http://apps.ideaconsult.net:8080/ambit2/dataset/54
Merge descriptor with activity - dataset with a web address again


http://apps.ideaconsult.net:8080/ambit2/dataset/54

Uniform approach to models creation

Read data from a web address - process - write to a web address

Dataset
- GET
- POST
- PUT
- DELETE

Feature
- GET
- POST
- PUT
- DELETE

Compound
- GET
- POST
- PUT
- DELETE

Algorithm
- GET
- POST
- PUT
- DELETE

Model
- GET
- POST
- PUT
- DELETE

http://myhost.com/dataset/trainingset1
http://myhost.com/algorithm/neuralnetwork
http://myhost.com/model/predictivemodel1

OpenTox
March 17, 2011
Ideaconslt Ltd.
Build a regression model

- http://apps.ideaconsult.net:8080/ambit2/algorithmlLR

- Clicking ‘Predict’ button initiates calculations
- It actually sends HTTP POST command to the algorithm URL, with parameter the dataset URL
- The demo dataset web address
- http://apps.ideaconsult.net:8080/ambit2/dataset/54
The model

Web address of the model could be retrieved via 
http://apps.ideaconsult.net:8080/ambit2/model/57/predicted, in this case it is the web address 

Web address of X variables could be retrieved via 
http://apps.ideaconsult.net:8080/ambit2/model/57/independent, in this case it is the web address 

--- WEKA Model ---
structures required NO
-- Data header --
attribute http://apps.ideaconsult.net:8080/ambit2/feature/22200 numeric
attribute http://apps.ideaconsult.net:8080/ambit2/feature/22213 numeric
@data
-- Class Index 0--
attribute http://apps.ideaconsult.net:8080/ambit2/feature/22200 numeric
Linear Regression Model
http://apps.ideaconsult.net:8080/ambit2/feature/22200 =
-0.2593 * http://apps.ideaconsult.net:8080/ambit2/feature/22213 +
-4.4086
Uniform access to calculations

Read data from a web address - process - write to a web address

Dataset
GET
POST
PUT
DELETE

Feature
GET
POST
PUT
DELETE

Compound
GET
POST
PUT
DELETE

Model
GET
POST
PUT
DELETE

Dataset
GET
POST
PUT
DELETE

Predicted results

http://myhost.com/dataset/newcompounds

http://myhost.com/model/predictivemodell

http://myhost.com/dataset/predictedresults1

Ideaconsult Ltd.
March 17, 2011
Apply the new regression model

- http://apps.ideaconsult.net:8080/ambit2/model/57 (Model)

- The demo dataset web address
Predictions, finally (at a dataset web address)

Advanced options
(a bit of command line, no web browser)

- [http://curl.haxx.se/](http://curl.haxx.se/)
- curl is a command line tool for transferring data with URL syntax
- Copy the dataset to a new one:
  ```
curl -X POST -d "dataset_uri=http://apps.ideaconsult.net%3A8080%2Fambit2%2Fdataset%2F54%3Ffeature_uri%5B%5D%3Dhttp%3A%2F%2Fapps.ideaconsult.net%3A8080%2Fambit2%2Ffeature%2F22213%26feature_uri%5B%5D%3Dhttp%3A%2F%2Fapps.ideaconsult.net%3A8080%2Fambit2%2Ffeature%2F22200%26feature_uri%5B%5D%3Dhttp%3A%252F%252Fapps.ideaconsult.net%253A8080%252Fambit2%252Fmodel%252F57%252Fpredicted"
http://apps.ideaconsult.net:8080/ambit2/dataset
```

The new URL returned is
Uniform approach to models validation and report generation

Read data from a web address - process - write to a web address

- Dataset
  - GET
  - POST
  - PUT
  - DELETE

- Model generating predictions
  - GET
  - POST
  - PUT
  - DELETE

- Validation
  - GET
  - POST
  - PUT
  - DELETE

= Validation report

- Report
  - GET
  - POST
  - PUT
  - DELETE

http://myhost.com/dataset/trainingset1
http://myhost.com/dataset/predictedresults1
http://myhost.com/model/predictivemodel1
http://myhost.com/validation
http://myhost.com/report/1
Ask the Validation service at Freiburg to validate the model


Task URL returned [http://opentox.informatik.uni-freiburg.de/task/4062](http://opentox.informatik.uni-freiburg.de/task/4062)

Query if the task is completed:

- `curl -H "Accept:text/uri-list" http://opentox.informatik.uni-freiburg.de/task/4062`
- Yes, it is, here is the address of the validation object

[http://opentox.informatik.uni-freiburg.de/validation/18920](http://opentox.informatik.uni-freiburg.de/validation/18920)
The validation resource

curl http://opentox.informatik.uni-freiburg.de/validation/18920

---

:regression_statistics:
  :sum_squared_error: 66.4592554978574
  :sample_correlation_coefficient: 0.42022542118295
  :root_mean_squared_error: 0.929035636699631
  :mean_absolute_error: 0.752657632127617
  :r_square: 0.176589404608387
  :target_variance_actual: 1.06200222895221
  :target_variance_predicted: 0.187523074000816
  :date: 2010-08-03 09:41:48 +02:00
  :test_target_dataset_uri:
  :real_runtime: 1
  :prediction_dataset_uri:
  :percent_without_class: 0.0
  :num_without_class: 0
  :percent_unpredicted: 0.0
  :num_instances: 77
  :training_dataset_uri:
  :num_unpredicted: 0
The validation report

curl -d validation_uris=http://opentox.informatik.uni-freiburg.de/validation/18920 http://opentox.informatik.uni-freiburg.de/validation/report/validation http://opentox.informatik.uni-freiburg.de/task/4063 <---- Task URI

curl http://opentox.informatik.uni-freiburg.de/task/4063 ← Ask if completed

:description:
:date: 2010-08-03T09:48:31+02:00
:uri: http://opentox.informatik.uni-freiburg.de/task/4063
:hasStatus: Completed
:title:
:percentageCompleted: 0.0
:resultURI: http://opentox.informatik.uni-freiburg.de/validation/report/validation/160
:due_to_time: 2010-08-03T10:48:31+02:00
:creator:
The validation report (human readable this time)
http://opentox.informatik.uni-freiburg.de/validation/report/validation/160
OpenTox web services implementation

- Could be hosted on a single computer (even a laptop or netbook would do) or
- Could be transparently distributed on multiple servers in various physical locations, in particular for better reliability, resilience and scalability;
- OpenTox webservices could be deployed behind firewalls, in Intranets (or even offline), when very tight security policies would have to be met;
- Third parties, willing to deploy OpenTox webservices could select a relevant subset of services to run, tailored to their specific needs;
Make the model available

Register at OpenTox ontology service
- RDF tripple storage
- Accepts HTTP POST
- SPARQL endpoint

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

Becomes visible for applications
OpenTox is a framework that provides unified access to toxicity data, QSAR models, validation support, and interpretation aids. It is open source, allowing for contributions from toxicologists, QSAR modelers, and API development for new QSAR algorithms.

- **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

OpenTox services can be used to develop specific applications or embedded in workflow systems.
Two end user oriented demo applications, making use of OpenTox webservices, have been developed, deployed and are available for testing - [http://toxcreate.org](http://toxcreate.org) and [http://toxpredict.org](http://toxpredict.org);

- ToxCreate creates models from user supplied datasets;
- ToxPredict uses existing OpenTox models to estimate chemical compound properties.
**Exercises (4 options)**

http://www.ideaconsult.net/downloads/echeminfo/

- Interested in applying predictive models, but not in OpenTox implementation details
- Interested in building predictive models, but not in OpenTox implementation details
- End users, interested in learning implementation details of OpenTox web services.
- Developers
Thank you!