



Deliverable D5.2

Automated validation report generation

| | |
|-----------------|--|
| Grant Agreement | Health-F5-2008-200787 |
| Acronym | OpenTox |
| Name | An Open Source Predictive Toxicology Framework |
| Coordinator | Douglas Connect |



| | | |
|----------------------|--|----------------------------------|
| Contract No. | Health-F5-2008-200787 | |
| Document Type: | Deliverable Report | |
| WP/Task: | WP5 | |
| Name | Automated validation report generation | |
| Document ID: | OpenTox Deliverable Report 5.2 | |
| Date: | Oct 31, 2010 | |
| Status: | Final Version | |
| Organization: | DC/ALU | |
| Contributors | Martin Gütlein Dmitry Druzhilovsky Pantelis Sopasakis Andreas Karwath Barry Hardy | ALU IBMC NTUA ALU DC |
| Distribution: | Public | |
| Purpose of Document: | To document results for this deliverable | |
| Document History: | 1 - First Draft v0.1 on Sep 24 2 - Introduction and Summary added by IBMC on Oct 05 3 - QPRF section added by NTUA 4 - Introduction and Summary changed by ALU on Oct 06 5 - Editing by DC Oct | |

Table of Contents

| | |
|--|-------------------------------------|
| Acknowledgements | 4 |
| Summary | 5 |
| 1. Introduction | 6 |
| 2. REACH reporting within the OpenTox framework | 7 |
| 3. REACH report web service | 7 |
| 3.1 REACH report service API | 9 |
| 3.2 Curl examples for working with QMRF reports | 10 |
| 3.2.1 Create a QMRF report..... | 11 |
| 3.2.2 Get report in QMRF–xml format | 11 |
| 3.2.3 Update a QMRF report..... | 11 |
| 3.2.4 Start the QMRF editor with a QMRF report | 12 |
| 4. REACH report editors | 12 |
| 4.1 QMRF Editor | 12 |
| 4.2 QPRF Editor | 13 |
| 5. Discussion | 19 |
| 5.1 Further Working Directions | 19 |
| 5.2 Summary | Error! Bookmark not defined. |

Acknowledgements

Research Funding

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).

Project Partners

Douglas Connect (DC), In Silico Toxicology (IST), Ideaconsult (IDEA), Istituto Superiore di Sanita' (ISS), Technical University of Munich (TUM), Albert Ludwigs University Freiburg (ALU), National Technical University of Athens (NTUA), David Gallagher (DG), Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences (IBMC), Seascope Learning (SL), Jawaharlal Nehru University (JNU), Fraunhofer Institute for Toxicology & Experimental Medicine (ITEM).

Advisory Board

European Centre for the Validation of Alternative Methods, European Joint Research Centre, U.S Environmental Protection Agency, U.S. Food & Drug Administration, Nestlé, Roche, AstraZeneca, Lhasa, Leadscope, University of North Carolina, Pharmatropé, Bioclipse, EC Environment Directorate General, Organisation for Economic Co–operation & Development, CADASTER, Bayer Healthcare.

Correspondence

Dr. Barry Hardy, OpenTox Project Coordinator and Director, Community of Practice & Research Activities, Douglas Connect, Baermeggenweg 14, 4314 Zeiningen, Switzerland

Email: [barry.hardy –\(at\)– douglasconnect.com](mailto:barry.hardy-(at)-douglasconnect.com)

Summary

OpenTox is supporting the development of reporting capabilities for the generation and presentation of results of alternative testing methods including validation and reporting results of relevance to REACH¹. The OpenTox report generating component generates reports to present the results of predictions and (Quantitative) Structure Activity Relationship ((Q)SAR) model validations to the user in a structured reporting format. OpenTox reporting formats are guided by standards, templates such as (Q)SAR Model Reporting Format (QMRF) and the (Q)SAR Prediction Reporting Format (QPRF)², and OECD validation principles, which specify that to facilitate the consideration of acceptance of a (Q)SAR model for regulatory purposes, it needs to be associated with the OECD Guidelines for (Q)SAR Validation³.

This report describes and documents the initial progress which has been achieved within OpenTox with respect to the creation of automated reporting facilities for validated (Q)SAR models of predictive toxicology data. Within the OpenTox framework, we have defined an Application Programming Interface (API) for the (Q)SAR reporting web service as part of the OpenTox validation API (opentox.org/dev/apis). We have established web services to automatically generate QMRF reports, which summarize key information on (Q)SAR-based predictive toxicology models, including the results of any validation studies, structured according to the OECD (Q)SAR validation principles. The user may subsequently edit the reports using a QMRF Editor. OpenTox services have also been developed for the generation of QPRF reports for validated toxicology predictions. A QPRF Editor for the manual annotation of automatically generated QPRF reports has been developed. Both editors are implemented as standalone applications that can also be started with a web browser. The actual creation of the reports uses a separate OpenTox web service running in the background.

The initial OpenTox reporting web services support rapid generation of REACH relevant documents in the form of standardized reports for (Q)SAR-based predictive toxicology models and their predictions. The reporting web services allow for the downloading and uploading of reports which may be saved locally or online, enabling the collaborative editing and sharing of reports. The reporting services should allow all current and future OpenTox applications to easily create and access reports on validated predictive toxicology models, which with the addition of authorization and authentication can include the use of confidential resources.

¹ guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?vers=20_08_08

² tcsweb3.jrc.it

³ www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html

1. Introduction

OpenTox is supporting the development of reporting capabilities for the generation and presentation of results of alternative testing methods including validation and reporting results of relevance to REACH⁴. The OpenTox report generating component generates reports to present the results of predictions and (Quantitative) Structure Activity Relationship ((Q)SAR) model validations to the user in a structured reporting format. OpenTox reporting formats are guided by standards, templates such as (Q)SAR Model Reporting Format (QMRF) and the (Q)SAR Prediction Reporting Format (QPRF)⁵, and OECD validation principles, which specify that to facilitate the consideration of acceptance of a (Q)SAR model for regulatory purposes, it needs to be associated with the OECD Guidelines for (Q)SAR Validation⁶.

This report describes and documents the initial progress which has been achieved within OpenTox with respect to the creation of automated reporting facilities for validated (Q)SAR models of predictive toxicology data. To report results from (Q)SAR predictions or (Q)SAR models, the European Commission Joint Research Center (EC JRC) has issued the two QMRF and QPRF harmonised templates. Developers and end-users of (Q)SAR models can submit information on (Q)SAR models by using the QMRF format, which summarizes key information on (Q)SAR models, including the results of any validation studies, structured according to the OECD (Q)SAR validation principles⁷. Within the OpenTox framework, we have established web services to automatically generate such QMRF reports, which the user may subsequently edit, using a modified version of the QMRF Editor published by the EC JRC. The QPRF is a template to summarize and report substance-specific predictions generated by (Q)SAR models, which are represented by QMRF. OpenTox services have been developed for the generation of QPRF reports for validated predictions. We are currently developing a QPRF Editor for the manual annotation of automatically generated QPRF reports.

⁴guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf?vers=20_08_08

⁵ tcsweb3.jrc.it

⁶ www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html

⁷ www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html

2. (Q)SAR reporting for REACH

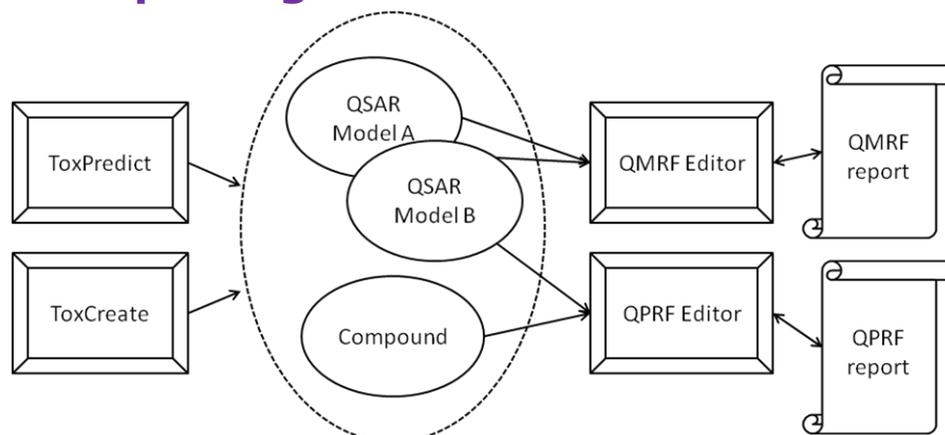


Figure 1 User perspective on (Q)SAR reports for REACH

Figure 1 shows the user perspective on working with QMRF and QPRF reports within the OpenTox framework. The workflow for reports allows OpenTox applications such as ToxCreate⁸ and ToxPredict⁹ to easily create and access reports. The reports can be created directly from the respective resources the user is working with within the applications: QMRF reports are created from models, QPRF reports are created from predictions, i.e. from a combination of models and compounds. For example, in ToxCreate a QMRF report is created automatically after a new model is built and validated. The user can edit, save, and export this report with the QMRF editor. Similarly, a QPRF editor is available for QPRF reports¹⁰. Both editors are implemented as standalone applications that can be started with a web browser. The actual creation of the report is done with a separate OpenTox web service running in the background.

3. (Q)SAR reporting web service for REACH

An OpenTox (Q)SAR reporting web service was designed to manage QMRF and QPRF reports for REACH submission purposes. The initial implementation of the service is available at opentox.informatik.uni-freiburg.de/validation/reach_report.

⁸ www.toxcreate.net

⁹ www.toxpredict.net

¹⁰ It is planned that a QPRF report can automatically be created and for a prediction performed in ToxPredict.

Figure 2 shows how the creation of a report works. A QMRF report is created from an existing (Q)SAR model that is provided to the web service as a URI parameter. The web service internally collects information from a range of other web services to automatically fill in the report content. For example, it queries the validation web service to add all cross validations that have been performed for the algorithm and training dataset (that have been used for building the model). The created QMRF report is stored at the report service. When creating a QPRF report, the compounds which are predicted by the model are required as additional input parameters¹¹.

Like all OpenTox resources, each report is identified and can be accessed via its URI. The report is made available in the official xml format¹², as well as in RDF xml (which is the common data exchange format within the OpenTox framework).

The web service furthermore allows the user to update and delete existing reports. The following sections describe the web service functionality in more details: the Application Programming Interface (API) definition for the service is presented in section 0, examples on how to use the API via the command line cURL tool are given in section 0.

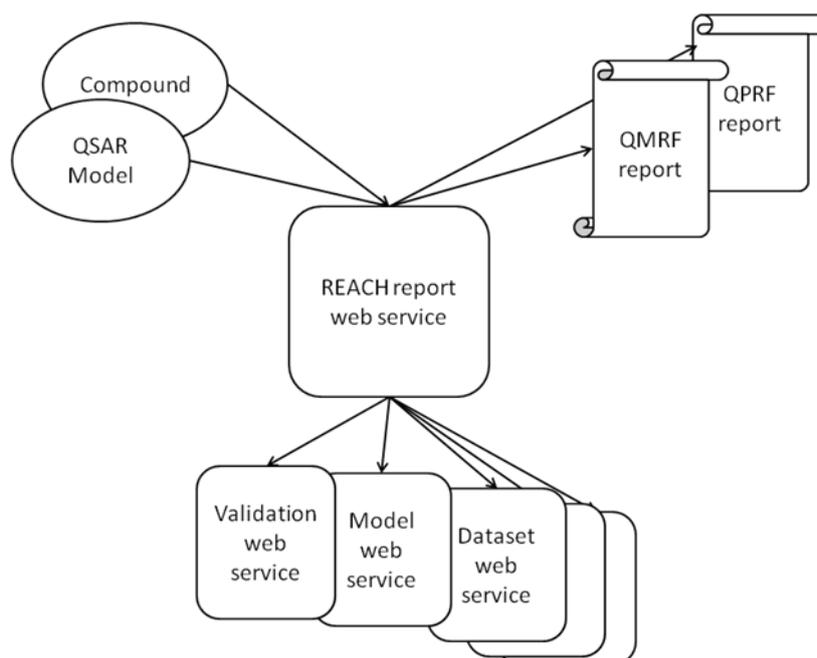


Figure 2 A web service for creating reports

¹¹ The QPRF service implementation is an ongoing development

¹² The Document Type Definition (DTD) for the QMRF xml can be found at ambit.sourceforge.net/qmrf/qmrf.dtd; an official xml format for QPRF has yet to be defined.

3.1 (Q)SAR REACH reporting service API

The API for the (Q)SAR REACH reporting web service is provided in Table 1. It is part of the validation API and can be found on the OpenTox web pages as the current version 1.1 (opentox.org/dev/apis/api-1.1/Validation) and the next version 1.2 scheduled for release in late 2010 (opentox.org/dev/apis/api-1.2/Validation).

| Description | Method | URI | Parameters | Result | Status codes |
|--------------------------------------|--------|--------------------------------------|--|--|-----------------|
| Create QMRF report | POST | /reach_report/qmrf | application/x-form-www-urlencoded model_uri =Model URI or application/qmrf-xml for creating a report with predefined QMRF XML content | Report URI or Task URI | 200,400,404,500 |
| Replaces QMRF report | POST | /reach_report/qmrf/{reportid} | application/qmrf-xml for creating a report with predefined QMRF XML content | Report URI | 200,400,404,500 |
| Update partially QMRF report | PUT | /reach_report/qmrf/{reportid} | validation_uri = a List of crossvalidation URIs and/or validation URIs of the same model < report_section (as defined in qmrf.dtd)> = content as string | Report URI or Task URI | 200,400,404,500 |
| Delete QMRF report | DELETE | /reach_report/qmrf/{reportid} | deletes the report | | |
| Retrieves the report | GET | /reach_report/qmrf/{reportid} | retrieves the report | representation, , format specified by MIME type (XML, RDF, HTML, PDF, XLS, where applicable) | |
| Start qmrf editor with report | GET | /reach_report/qmrf/{reportid}/editor | - | return jnlp, starts QMRF editor as Java webstart application | 200,404 |
| Create QPRF | POST | /reach_report/ | application/x-form-www-urlencoded | Report URI or Task URI | 200,400, |

| | | | | |
|-------------------------------------|--------|-------------------------------|---|--|
| report | | qprf | | 404,500 |
| | | | model_uri = Model URI One of { dataset_uri = Dataset URI compound_uri = compound uri }, specifying the compounds or application/qprf-format-to-be-defined for creating a report with predefined QPRF content | |
| Replaces QPRF report | POST | /reach_report/qprf/{reportid} | same as above, replaces the content | Report URI |
| Updates QPRF report | PUT | /reach_report/qprf/{reportid} | same as above, but adds new content to the report | Report URI |
| Deletes QPRF report | DELETE | /reach_report/qprf/{reportid} | deletes the report | |
| Retrieves the report content | GET | /reach_report/qprf/{reportid} | retrieves the report | representation, , format specified by MIME type (XML, RDF, HTML, PDF, XLS, where applicable) |
| Report searching facilities | GET | /reach_report/{type} | application/x-form-www-urlencoded any or subset of model_uri = Model URI dataset_uri = Dataset URI compound_uri = Compound URI algorithm_uri = Algorithm URI endpoint_uri = endpoint URI, as defined by the ontology search =any free text, etc. | Retrieves list of reports, related to the model, specified by any of the parameter URI |

Table 1 The REACH report web service API

3.2 cURL examples for working with QMRf reports

In this section we will examine how to access the validation web services using the command line cURL tool (curl.haxx.se) which can be used to formulate and execute API-compliant commands.

Alternatively, the REST calls can be performed with an arbitrary REST client and/or most major modern programming languages of relevance.

3.2.1 Create a QMRF report

A QMRF report can be created in two ways:

- a) The standard way is to specify a model URI as parameter:

```
curl -X POST -d model_uri=http://opentox.informatik.uni-freiburg.de/model/1
http://opentox.informatik.uni-freiburg.de/validation/reach_report/QMRF
```

This returns a task-object as result¹³. As soon as the task is completed, i.e. the report generation is finished, the QMRF report URI is set as resultURI property of the task object.

- b) Alternatively, the user can create a QMRF report resource on the web service from a QMRF report that already has been created externally (e.g. with the QMRF editor). In this case the report has to be uploaded to the service in QMRF-xml format:

```
curl -X POST -T QMRF_v1.2_FishTox.xml http://opentox.informatik.uni-
freiburg.de/validation/reach_report/QMRF
```

(The above QMRF example report file is located at ecb.jrc.ec.europa.eu/qsar/qsar-tools/qrf/QMRF_v1.2_FishTox.xml)

3.2.2 Get report in QMRF-xml format

After creating a report, the web service provides the report URI as a result. This URI identifies the report and can be used to download the report from the service. Set the accept-header of the REST call to "application/qmrf-xml" in order to derive the report in the official QMRF-xml format.

```
curl -H "accept:application/qmrf-xml" http://opentox.informatik.uni-
freiburg.de/validation/reach_report/QMRF/3
```

3.2.3 Update a QMRF report

To update an existing QMRF report, the user can edit the xml, and post it to the URI of the existing report:

```
curl -X POST -T QMRF_v1.2_FishTox.xml http://opentox.informatik.uni-
freiburg.de/validation/reach_report/QMRF/3
```

¹³ See opentox.org/dev/apis/api-1.1/AsyncTask

3.2.4 Start the QMRF editor with a QMRF report

A QMRF report can be edited with the QMRF editor (see section 0). To this end, open the following link in a browser (with Java 1.6 Webstart enabled):

http://opentox.informatik.uni-freiburg.de/validation/reach_report/QMRF/3/editor

4. (Q)SAR reporting editors for REACH

The complete content of QMRF and QPRF reports cannot be generated fully automatically. There are some fields that require user input, e.g., the mechanistic interpretation of the model (if possible) in the QMRF report as required by the fifth OECD Validation Principle. To this end, OpenTox is providing two editors to work with the reports. Figure 3 visualizes that both, the QMRF editor and the QPRF editor, can be used in a flexible way. They can load, edit and store reports to/from the REACH reporting web service (introduced in section 3), as well as to the local file system of the user. Furthermore, it is possible to export reports in PDF format. The following sub-sections here introduce both reporting editors in more detail.

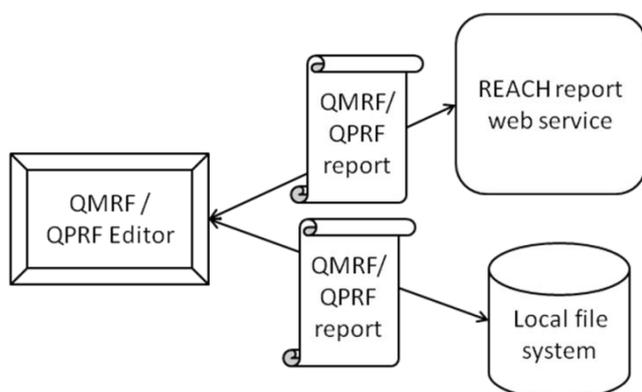


Figure 3 QMRF and QPRF Editors can be used to edit/store/export reports.

4.1 QMRF Editor

The original QMRF editor was developed by the OpenTox partner IDEA¹⁴. It is an open source Java application, and can be started as a Web Start application¹⁵. The original functionality allows creating a new report from scratch. It is further possible to load existing reports that are stored in the predefined QMRF-xml format. Each section of the report can be edited via text fields or forms

¹⁴ See ambit.acad.bg

¹⁵ ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=QRF

that provide more guidance (i.e. for QMRF authors). Help dialogs are available for every section. QMRF reports can be stored locally in QMRF–xml format, and can be exported to PDF. This QMRF editor has been extended to meet the new requirements within the OpenTox framework. As described in section 3.2.4, the QMRF editor will start and directly download the respective QMRF report when adding the suffix ‘/editor’ to the QMRF report URI.¹⁶ It is further possible to manually download another report from the web service. Figure 4 shows a screenshot of the new QMRF editor when manually downloading a report. Moreover, the user can upload changes to the web service, by either overwriting the existing report, or creating a new report on the server.¹⁷

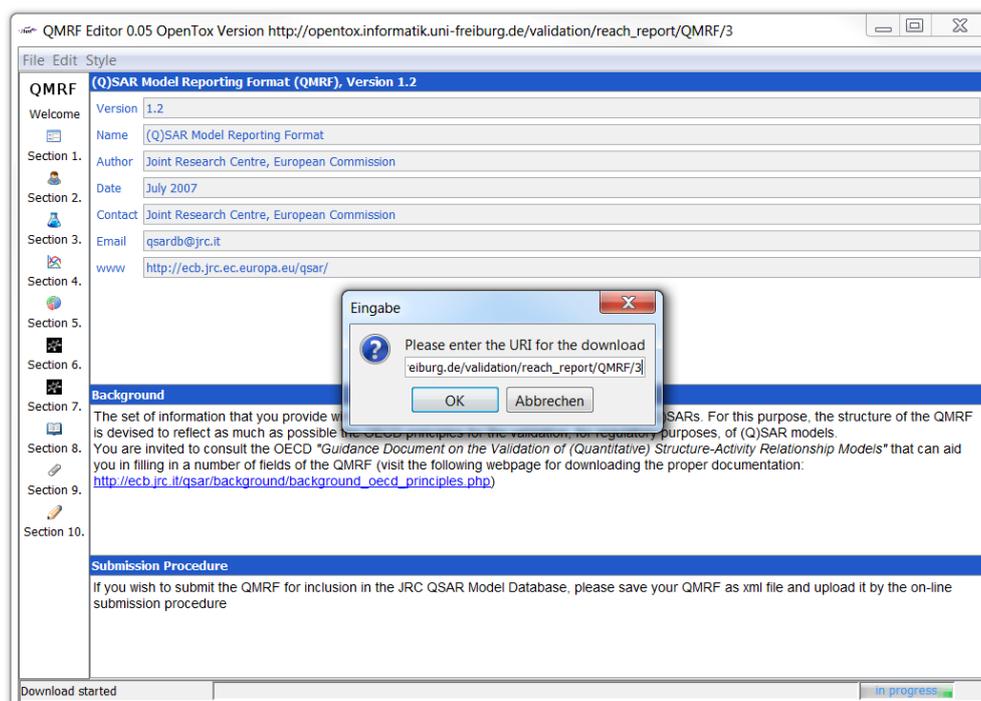


Figure 4 Download a QMRF report with the QMRF editor

4.2 QPRF Editor

Q–edit is a new QPRF editor developed under OpenTox which aims at exploiting implemented web services to provide functionalities that facilitate the creation of QPRF reports by an end user.

Briefly, the main use case consists of the following steps:

- a) Create a new (empty) QPRF report (see Figure 5).

¹⁶ Open opentox.informatik.uni-freiburg.de/validation/reach_report/QMRF/3/editor with a Java Web Start–enabled browser.

¹⁷ The next version of the QMRF editor plans to support Authorization & Authentication

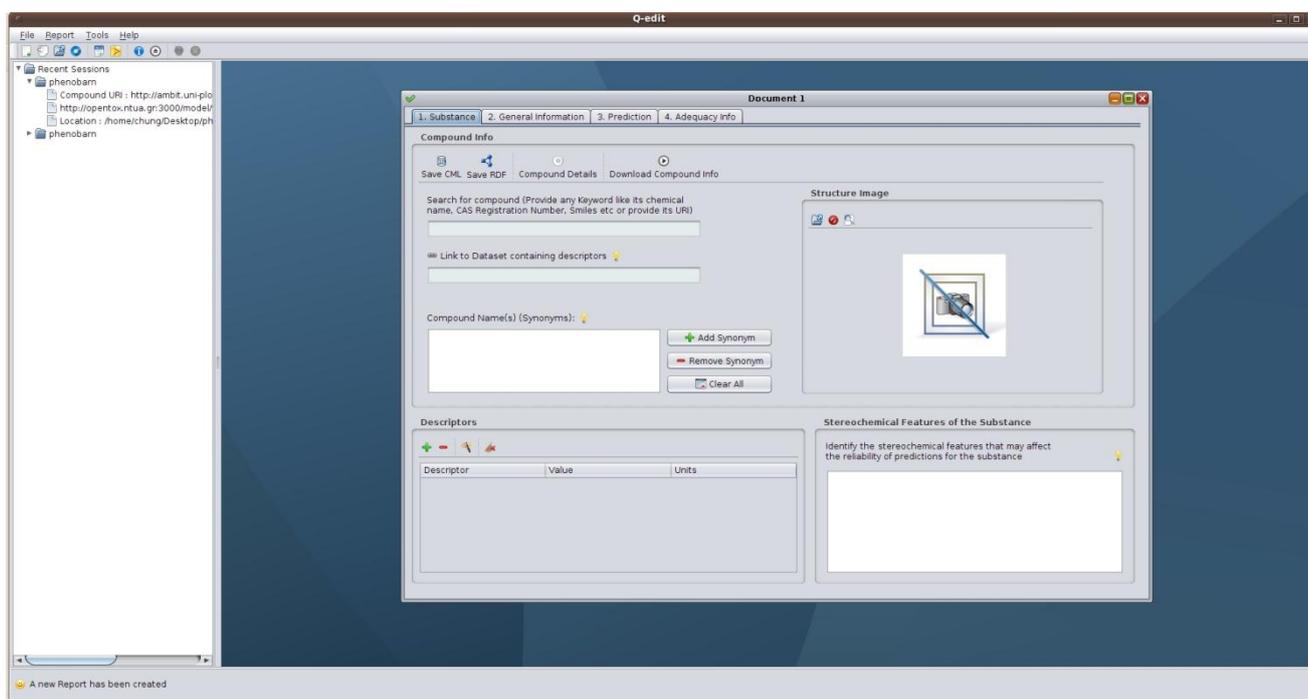


Figure 5 Create a new report

- b) Authenticate yourself providing your credentials. Optional Step, users will be prompted to login in case they try to access some restricted resource. (See Figure 6).



Figure 6 Users provide their credentials and initiate a new authentication session.

- c) Search for a compound in an on-line database (e.g. AMBIT) - Inspect the downloaded compound (View Chemical name(s), SMILES string, CAS RN and a depiction of the compound). Enter additional meta information about the compound, e.g. discuss its stereochemical features that might affect the validity of the prediction. (see Figure 7)

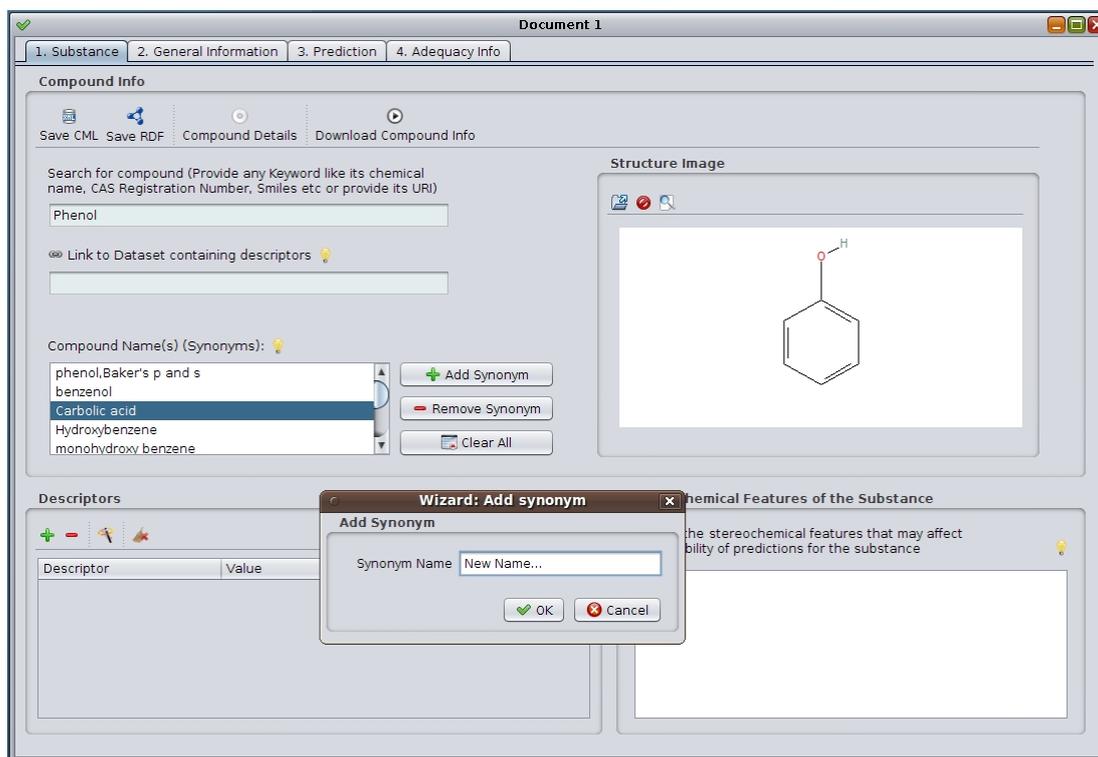
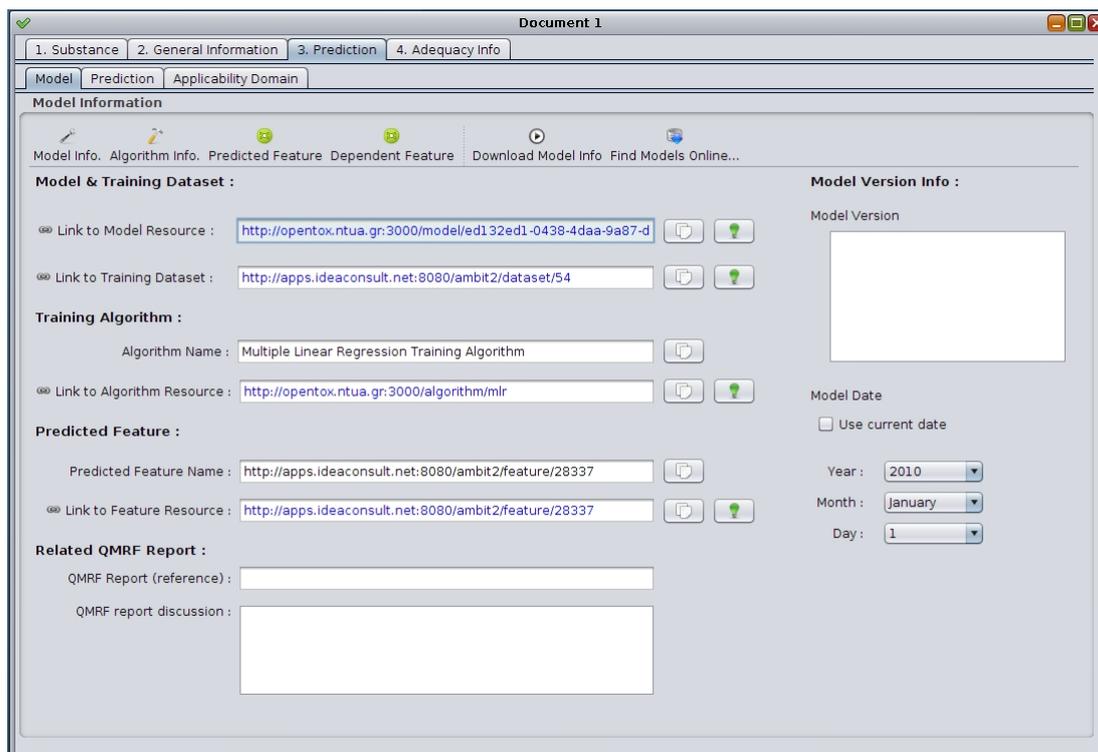


Figure 7 Information about Phenol is loaded from a remote service

- d) Select a model from a list or look up for a model according to some search criteria (e.g. some endpoint) – Inspect the downloaded model (Get a list of all independent features for the model); access information concerning the dependent and predicted feature; examine the training algorithm and get further (meta)information about it; find QMRF reports created for that model; add some discussion about the model. (see Figure 8).



The screenshot shows the 'Document 1' window in the OpenTox application. The interface is divided into several sections:

- Model Information:** Includes tabs for 'Model', 'Prediction', and 'Applicability Domain'. Below these are sub-tabs: 'Model Info.', 'Algorithm Info.', 'Predicted Feature', 'Dependent Feature', 'Download Model Info', and 'Find Models Online...'. A toolbar contains icons for a pencil, eraser, smiley face, sad face, play button, and globe.
- Model & Training Dataset:**
 - 'Link to Model Resource': <http://opentox.ntua.gr:3000/model/ed132ed1-0438-4daa-9a87-d>
 - 'Link to Training Dataset': <http://apps.ideaconsult.net:8080/ambit2/dataset/54>
- Training Algorithm:**
 - 'Algorithm Name': Multiple Linear Regression Training Algorithm
 - 'Link to Algorithm Resource': <http://opentox.ntua.gr:3000/algorithm/mlr>
- Predicted Feature:**
 - 'Predicted Feature Name': <http://apps.ideaconsult.net:8080/ambit2/feature/28337>
 - 'Link to Feature Resource': <http://apps.ideaconsult.net:8080/ambit2/feature/28337>
- Model Version Info:**
 - 'Model Version': (Empty text area)
 - 'Model Date':
 - Use current date
 - Year: 2010
 - Month: January
 - Day: 1
- Related QMRF Report:**
 - 'QMRF Report (reference)': (Empty text area)
 - 'QMRF report discussion': (Empty text area)

Figure 8 Information about a model of given URI is loaded from a remote location

- e) Use the model to obtain a prediction for the compound (Under development for next version; for now users have to enter the predicted value manually). Write some comments about the predicted value.
- f) Acquire a list of structural analogues of the compound for which the QPRF report is created and append some discussion. (see Figure 9)

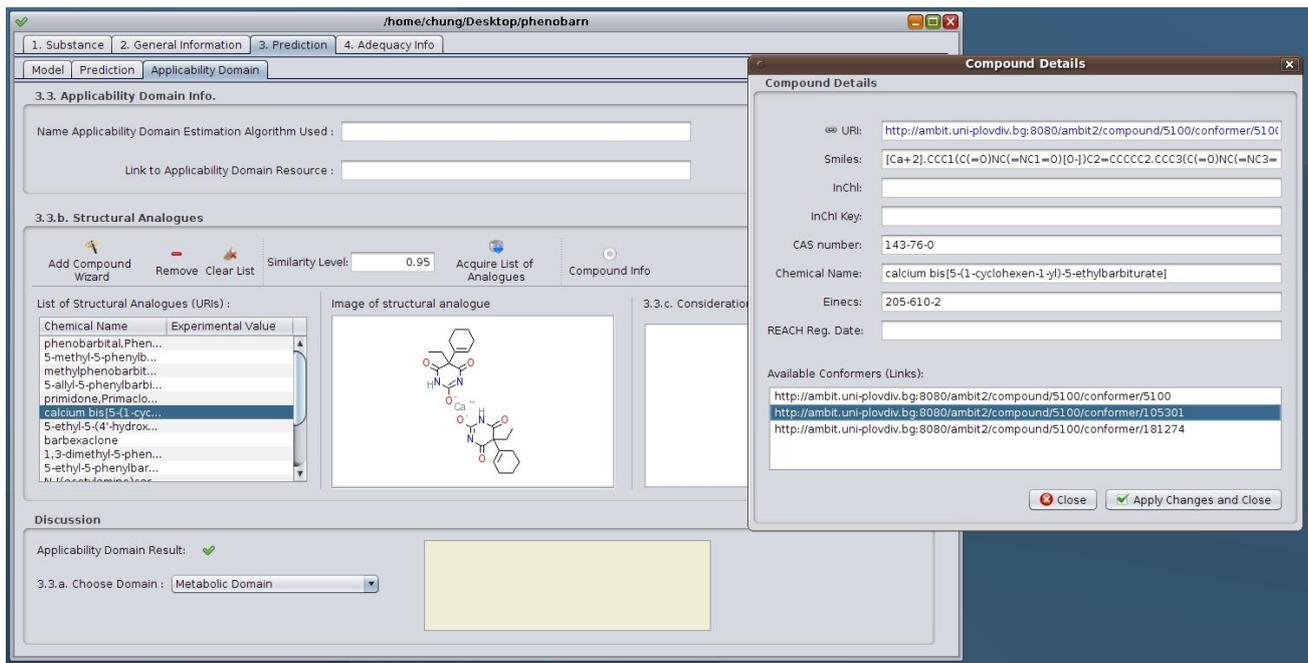


Figure 9 List of structural analogues of the compound PHENOBARN. For each compound users have access to some of its basic characteristics.

g) Add information about Authors of the QPRF report. (See Figure 10)

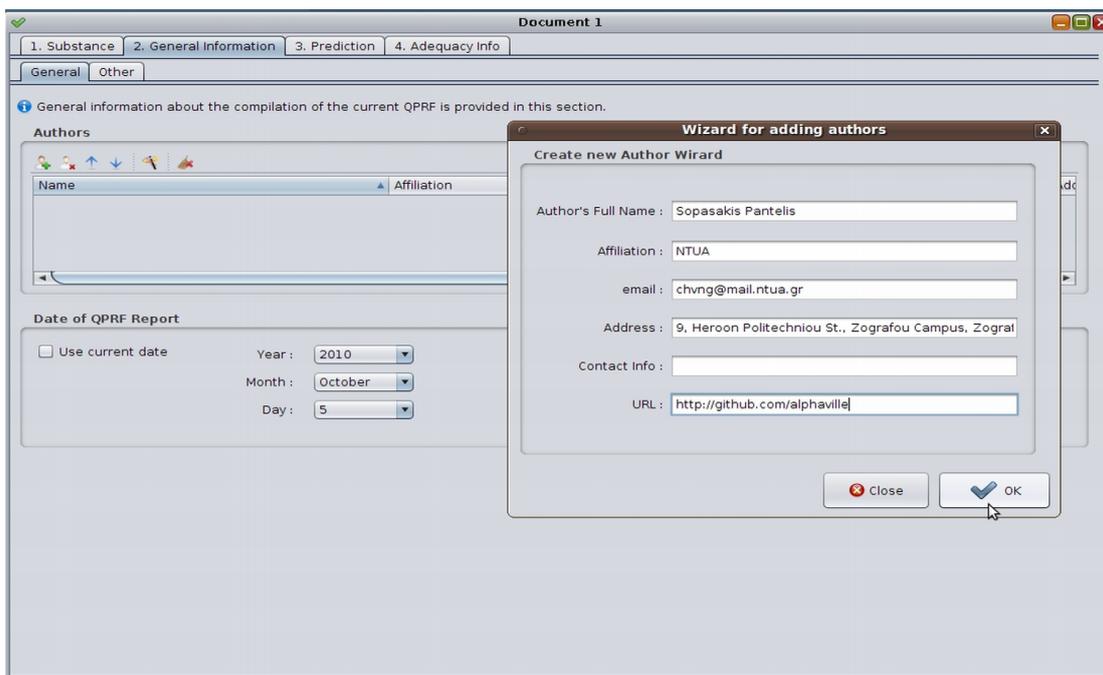


Figure 10 Users provide personal information and contact details

- h) Export the report in PDF format. The resulting document is fully compliant with the standards for QPRF reports that are provided by the EC JRC¹⁸. Export in RTF format is under development.
- i) Configure the editor. (see Figure 11)



Figure 11 Configuration Dialog Box – Editor Options

- j) Save the report in binary format in a local directory. (Loading from/Saving to remote servers is also under development.) Users can also lock a report with a passphrase. (see Figure 12).

¹⁸ ecb.jrc.ec.europa.eu/qsar/qsar-tools/qrf/QPRF_version_1%201_DEREK_SS.pdf

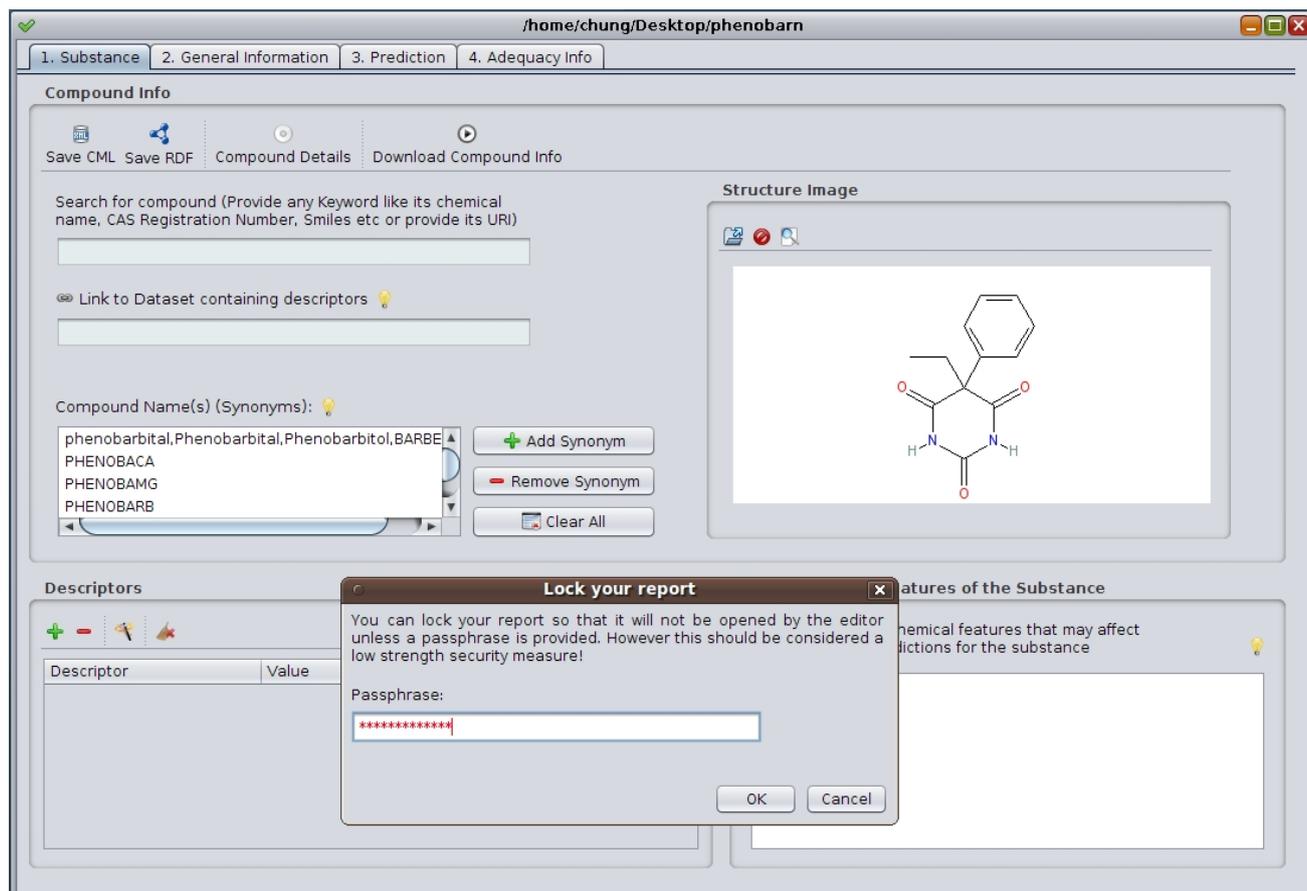


Figure 12 Protect your reports with a passphrase

Users are guided through the above steps with jargon-free documentation that map directly to the sections of the QPRF report as described by the EC JRC. Though it can be used in offline mode, Q-edit is designed to interact with various OpenTox web services providing real-time access to compound databases and model repositories. QPRF reports are serialized in a compressed binary format so that save/open operations are supported. However, for the sake of uniformity and transparency, QPRF reports are stored in RDF format.

Q-Edit is a tool that allows users to create new prediction reports, and to manage and inspect existing ones. The Q-edit application is written in Java using JDesktop, Swing and AWT and is licensed under the GNU GP License, v.3.0. The source code is available for download from github.com/alphaville/Q-edit and the executable can be downloaded from github.com/alphaville/Q-edit/downloads. It can also be compiled as a Java Web Start application.

5. Discussion

5.1 Further Working Directions

In the near future OpenTox reporting services for REACH purposes will be extended within the OpenTox framework by:

- Achieving improved and complete automatic generation of QMRF and QPRF reports (while allowing for user interaction for writing and editing where needed);
- Extension of the new QPRF editor, including its integration within the OpenTox Framework and relevant web services;
- Integration of authentication and authorization into the tools, allowing for confidential information to be integrated as well as allowing restricted access to certain reports.

5.2 Conclusions

The initial OpenTox reporting web services support rapid generation of REACH relevant documents in the form of standardized reports for (Q)SAR-based predictive toxicology models and their predictions. The reporting web services allow for the downloading and uploading of these reports which may be saved locally or online, enabling the collaborative editing and sharing of reports. Furthermore, the QMRF and QPRF editors support the manual editing of reports which is inevitably required for the addition of expert interpretations and clarifications.