

Principles and process for developing and donating models (QSAR models, profilers, calculators etc.) and databases to the QSAR Toolbox

1. Introduction

With the continuous increase of popularity of the OECD QSAR Toolbox, there are opportunities to extend the functionality of the QSAR Toolbox through additional QSAR models, profilers and databases. This is very welcomed because inclusion of additional QSAR models, profilers and databases increases the value of the Toolbox, however, adding these to the Toolbox presents technical challenges related to the following aspects:

- **Process for adding donated models/databases:** it is anticipated that there will be more interested in donating QSAR models, profilers and databases to the QSAR Toolbox, and there is a need to develop a process for technical implementation of these requests.
- **Resources:** financial support is needed to assure donated extensions are compatible with the QSAR Toolbox.

In discussions, it was noted that adding new models and databases directly within the Toolbox would require significant programming resources and increase the size of the Toolbox. The QSAR Toolbox Management Group discussed these aspects and possible solutions at the 14th meeting in November 2017. In order to address the above challenges, the API of the QSAR Toolbox was made available for plug-in development in the version 4.3, which was released in early 2019.

2. Proposed process for developing and donating new models and databases to the QSAR Toolbox.

The QSAR Toolbox Management Group suggested that new models and databases could be added to the Toolbox as plugin extensions. The processes for developing and donating new models (section 2.1) and databases (section 2.2) are found below. These processes were agreed by the QSAR Toolbox Management Group at the 14th meeting of the QSAR Toolbox Management Group in November 2017, and considering comments from the QSAR Toolbox Management Group on the teleconference in June 2018.

2.1. Process for developing and donating models (QSAR models, profilers, calculators, metabolism simulators and other models) as extensions to the Toolbox

Solutions were proposed at the November 2017 QSAR Toolbox Management Group for strategies to handle requests to add new models into the Toolbox in order to best focus resources, minimise technical issues and maintain the high quality of the Toolbox. The QSAR Toolbox Management Group agreed to an approach, where new models are added as extensions to the Toolbox after reviewing the quality and usefulness of new model proposed for inclusion.

2.1.1. Proposal for developing and donating new models

- a. A sponsor will be required for new models. The sponsor/model developer is required to provide the IT capabilities needed develop the extensions to add to the QSAR Toolbox.
- b. The developer will need to provide a description and associated background information of the model.
 - i. For QSAR models, the following documents should be provided:
 - A QMRF (QSAR Model Reporting Format) document for each model documenting how the QSAR models meet the OECD validation principles for QSARs.
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono\(2007\)2&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en)
 - ii. For new or revised profilers/metabolism simulators, the following documents should be provided:
 - A description of the profiler and associated background information using the templates for profilers (**Appendix 1**), which was approved at the 12th meeting of the QSAR Toolbox Management Group.
 - iii. For a parameter calculator, the following documents should be provided:
 - A description of the parameter calculator and associated background information
- c. The new model (QSAR models, profiler or metabolism simulator, calculators, etc.) should be presented to the QSAR Toolbox Management Group at a meeting or teleconference, however this will not constitute an endorsement step. This presentation could be done before or after the models is made available so as to not limit the addition of extensions to twice a year.

- d. If the sponsor/developer is a member of the QSAR Toolbox Management Group, the sponsor can inform the Management Group in advance that the sponsor is planning to develop the model as an extension and Management Group members can express their interest in providing input to the development. If the Toolbox Management Group members express interest in providing input on an model, the sponsor will coordinate input of interested members (e.g. form a project team).

2.1.2. Sponsor/model developer prepares the package which contains all the required files.

The package can be one of following two types:

- a. Software which includes the model (QSAR model, parameter calculator, profiler, or metabolism simulator) and a plugin allowing the model to dock to the Toolbox via the Toolbox application programming interface (API¹). Extensions that are docked to the Toolbox via the API will limit the increase in size of the Toolbox and streamline resources needed for making extensions available. However, this package type requires the developer to have IT capability.
 - i. In case the developer cannot provide a plugin module to implement the Toolbox API, a sponsor/developer will be required to do this for each model.
- b. A configuration which can be interpreted by the existing system modules. Supported configurations are:
 - ii. QSAR models defined by integrated editor², which needs to be used to define all of the model's metadata
 - 1. For calculation of target endpoint value there are available two scenarios:
 - a. The QSAR equation is clearly defined and all required parameters are available in the system. In this case, the equation needs to be entered using the QSAR editor.
 - b. Web services for calculation of predicted value are available. In this case, the address of the web services needs to be entered using the QSAR editor. The model developer is responsible for keeping his QSAR service up and running.
 - iii. Profilers created by integrated profiling editor
 - iv. Data file with pre-calculated (tested) values by which a custom calculator can be defined.

2.1.3. Confirmation of the package

- a. Levels of checks:

¹ Toolbox API implementation allows docking of QSARs, profilers, metabolism simulators and parameter calculators (<https://qsartoolbox.org/developers/api/>)

² The system already includes (from version 1.0) an editor for profiling schemes. By using this editor, the users can produce and exchange configurations for custom profiling schemes.

v. Technical check

1. package does not contain harmful code
2. package can be installed and recognized correctly by the system
3. the modules work as it is expected
4. module is not crashing the system

vi. Completeness check:

1. Contains all of the required documentations
- b. Confirmation without “Toolbox Repository” – the model developer provides the package to the QSAR Toolbox Coordination Group (ECHA/OECD/the Toolbox developer).
- c. Confirmation with “Toolbox Repository”³ – developer uploads the package to the Toolbox store for confirmation.
- d. For both b. and c., until the technical and completeness checks are completed to confirm that items have been provided and are functioning correctly, the package will be accessible only to the Toolbox Coordination Group. Separate confirmation emails are sent by the actors performing the technical and the completeness check to inform the Coordination Group and the model developer about the outcome of the checks.

2.1.4. Models are available to the public

- a. Once the confirmation steps are successfully completed, the model will be made available to users of the QSAR Toolbox as an extension on the Toolbox web site. These extensions will be available via the QSAR “Toolbox Repository”.
- vii. Without “Toolbox Repository” - The green-lighting group will need to provide the package to Toolbox website administrator who will need to prepare a link for the extension in the “download” section.
- viii. With “Toolbox Repository” - After successful confirmation the package will be accessible for download for any user of the system.
- b. Interested QSAR Toolbox users can install the extension via Toolbox repository.

2.2. Proposed process for developing and donating databases (other than those integrated into the Toolbox installation)

The process for donating databases is similar to the process for developing and donating models as extensions to the Toolbox. The standard process for donating databases that are integrated into the Toolbox installation is described below (section 2.2.1). In addition, new databases can be added to the Toolbox as extensions using the Toolbox import functions (section 2.2.2).

³ <https://repository.qsartoolbox.org/>

2.2.1. Continue standard approach for developing and donating databases that are integrated into the Toolbox installation

- a. All fields of the database must be well defined. The sponsor/database developer will provide a description of the database and associated background information using the template for documentation of databases (**Appendix 2**), which was approved at the 12th meeting of the QSAR Toolbox Management Group. The donor is also requested to provide a letter, which should explain the re-use policy of the data in the Toolbox and clarify the ownership of the data (including eventual intellectual property rights).
- b. The database should be presented to the QSAR Toolbox Management Group at the meeting or call and the group must endorse the inclusion of the database in the Toolbox. Endorsement can be done by written procedure in cases where timing does not align with the Toolbox Management Group meetings.
- c. The new database will be integrated in the installation package of the next version of the QSAR Toolbox.

2.2.2. Proposed approach for developing and donating databases as extensions of the Toolbox

- a. The database is to be documented using the database template (**Appendix 2**).
- b. In order to be made available as a QSAR Toolbox extension, the sponsor/database developer will need to:
 - i. Provide the excel file for the database.
 - ii. Provide a description of the database and background information using the template for the documentation databases (**Appendix 2**).
 - iii. Provide a letter, which should explain the re-use policy of the data in the Toolbox and clarify the ownership of the data (including eventual intellectual property rights).
- c. The database should be presented to the QSAR Toolbox Management Group at the meeting or call, however this will not constitute an endorsement step. This presentation could be done before or after the extension is made available so as to not limit the addition of extensions to twice a year.
- d. Once members of the Toolbox Coordination Group have confirmed completeness of the information, the database excel files will be made available to users of the QSAR Toolbox as import files on the Toolbox web site. Users can download and import the database to their QSAR Toolbox. Further distribution of databases will be through a QSAR Toolbox extension 'Repository' where these plugins will be available.
- e. The QSAR Toolbox users can make use of the database extensions that serve their purposes.

3. Updating of Extensions

As the extensions become available on the Toolbox Repository website, there are opportunities to take a collective approach to improve them. If a QSAR Toolbox user wishes to update or supplement an extension, this could be done using the existing extension as a starting point (this would not apply to plug-ins for locally installed QSAR modules). The new sponsor/developer could work independently or collaboratively with the original developer and provide, using the same process above, an updated version of the plug-in to be made available on the extension website.

4. Future Integration into QSAR Toolbox Download

If a particular extension becomes very popular with a wide cross-section of users after being made available on the extension website, the QSAR Toolbox Management Group could consider integrating the extension into the main QSAR Toolbox Download. This would follow an endorsement process by the QSAR Toolbox Management Group.

Appendix 1. Template for Models or Profilers

About section of a model/profiler	
Name of the model/profiler	
Sponsor; Developer; date; version	
<i>Sponsor:</i>	
<i>Developer:</i>	
<i>Version:</i> (Date)	
Relevance/Applicability to endpoint(s)	
Relevance/Applicability to particular chemical classes	
Approach used to develop the model/profiler - Concise but informative description of:	
a) The aim of this model/profiler is to	
b) The model/profiler was developed on a basis of	
c) The model/profiler was developed from a dataset of	
d) Reference source	
Summary description of <u>profiles/alerts</u> within the model/profiler	
Profile/structural alert	Phys-chem parameter
Counter category:	
Similar to other model/profilers	
Short description of update version	
Disclaimer	
The sponsor/model developer hereby grants a worldwide, non-exclusive, royalty-free, perpetual license to the OECD and their contractors, if any, to include and use the donated model as part of QSAR Toolbox and to grant a sub-license to the QSAR Toolbox users to use the donated model as	

part of QSAR Toolbox in accordance with the QSAR Toolbox End User Licence Agreement (as amended from time to time).

The sponsor/model developer warrants that it owns or has secured all right, title and interest to the donated model, including data contained therein, if any, and that the granting of the license to the OECD and its contractors, if any, herein does not breach any third party rights. The sponsor/model developer agrees to indemnify, defend and hold the OECD and ECHA, and its contractors, if any, harmless from and against any and all claims, liability and costs, in connection with or arising out of infringement of the rights of any third party in the donated models, including data contained therein, if any.

The sponsor/model developer shall be responsible for the accuracy and/or completeness of the donated model, including data contained therein, if any.

The OECD and ECHA shall not be responsible for the use of or inability to use the donated model by QSAR Toolbox users.

The sponsor/model developer takes responsibility for the veracity of the information above.

Detailed description of each category/alert/boundaries

Individual profile/alert/boundaries/other info applicable for defining categories within a model/profiler		
Name		
Type of profile/alert		
Description		
Mechanistic Domain		
Mechanistic alert		
Mechanism		
Set of chemicals used for profile development (local training set)		
Data/Knowledge used for profile development		
Profile/alert analysis	Profile/alert	Number chemicals analysed (Sensitisers/ Total)
References		

Appendix 2. Template for Databases

About section of a database		
Name of the database		
Sponsor; Developer; Date; Version		
Sponsor:		
Developer:		
Version:		
Date of implementation:		
Date of update:		
Short description		
Type of endpoints		
Endpoints covered		
Biological and exposure factors:		
Sources:		
Quality assurance		
QA criteria	Yes/No	Number
1. Database Layout assessment		
Can the DB be imported without loss of information?		
Are there any conflicts between data?		
Are the endpoint paths clearly stated?		
2. QA assessment of chemical ID		
Conflict records		
Resolved conflicts		
Mixture		
Dissociating products		
Inorganic		
Anionic		
Cationic		
High QA relation CAS/2D		
Moderate QA relation CAS/2D		
Low QA relation CAS/2D		
High QA relation 2D/Name		
Low QA relation 2D/Name		
3. Preprocessing of data		

Aggregation of data		
Similar to other databases		
Short description of update version		
Disclaimer		
The sponsor/database developer hereby grants a worldwide, non-exclusive, royalty-free, perpetual license to the OECD, and its contractors, if any, to include and use the donated database as part of QSAR Toolbox and to grant a sub-license to the QSAR Toolbox users to use the donated database as part of QSAR Toolbox in accordance with the QSAR Toolbox End User Licence Agreement (as amended from time to time). The sponsor/database developer warrants that it owns or has secured all right, title and interest to the donated database, including data contained therein, and that the granting of the license to the OECD and its contractors, if any, herein does not breach any third party rights. The sponsor/database developer agrees to indemnify, defend and hold the OECD and ECHA harmless from and against any and all claim, liability and costs, in connection with or arising out of infringement of the rights of any third party in the donated databases, including data contained therein. The sponsor/database developer shall be responsible for the accuracy and/or completeness of the donated database, including data contained therein. The OECD and ECHA shall not be responsible for the use of or inability to use the donated database by QSAR Toolbox users. The sponsor/database developer takes responsibility for the veracity of the information above.		