OECD (Q)SAR Toolbox v.4.4

Tutorial on how to Import/Export a custom database and Import/Export a database via IUCLID
Outlook

• Aim
• Keywords
• Definition of Database/Inventory
• Import
• Export
Aim

This is a step-by-step presentation designed to take the user of the Toolbox through the process of importing/exporting of custom databases/inventories.

Note: Please note that building of custom items (such as profilers, (Q)SAR models as well as importing of custom databases) is only enabled in single user mode. So, if your Toolbox is installed in multiuser mode, you will be not able to follow this tutorial.
Outlook

- Aim
- **Keywords**
- Definition of Database/Inventory
- Import
- Export
Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options (e.g. Profiling)

WORKFLOW – the use, in combination, of the different modules (e.g. prediction workflow: from input to report)

IMPORT – Toolbox allows to import a custom databases with experimental data. Import should follows a specific import layout (Excel type of IU format)

EXPORT – Toolbox allows to export data and prediction for chemicals available on data matrix in a format appropriate for Excel

ENDPOINT TREE – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

DATA MATRIX – Table reporting the chemical(s) and data (experimental results, profilers outcomes, predictions). Each chemical is in a different column and each data in a different row
Outlook

- Aim
- Keywords
- **Definition of Database and Inventory**
- Import
- Export
Definition of database and inventory

• **A database** is a collection of structures accompanied with experimental data.

• **An inventory** is a collection of structures without experimental data.
Outlook

• Aim
• Keywords
• Definition of Database and Inventory
• **Import:**
  • Import of database
    • Vertical import
    • Horizontal import
  • Import of inventory
  • Import via IUCLID
• Export Data matrix
Import of database: Types of import

- **Vertical import:**
  It is appropriate for a set of chemicals with consistent experimental data and the same supporting information (e.g. endpoint, test organism, test condition, author etc.).

- **Horizontal import:**
  It is appropriate for a set of chemicals with different types of experimental data accompanied with supporting information (endpoints, test condition, test organism, author etc.).
Outlook

• Aim
• Keywords
• Definition of Database and Inventory
• **Import:**
  • Import of database
    • Vertical import
    • Horizontal import
  • Import of inventory
  • Import via IUCLID
• Export Data matrix
Import of database: Vertical import layout

- Vertical import layout contains two main sections: substance information (1) and experimental data (2);
- Each column (2) defines data points for a single experiment.
- The imported file can be in **xlsx** or **tsv** format.

<table>
<thead>
<tr>
<th>CAS #</th>
<th>NAME</th>
<th>SMILES</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>...</th>
<th>Experiment M</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAS 1</td>
<td>NAME 1</td>
<td>SMILES 1</td>
<td>Value 1.1</td>
<td>Value 2.1</td>
<td>...</td>
<td>Value M.1</td>
</tr>
<tr>
<td>CAS 2</td>
<td>NAME 2</td>
<td>SMILES 2</td>
<td>Value 1.2</td>
<td>Value 2.2</td>
<td>...</td>
<td>Value M.2</td>
</tr>
<tr>
<td>CAS 3</td>
<td>NAME 3</td>
<td>SMILES 3</td>
<td>Value 1.3</td>
<td>Value 2.3</td>
<td>...</td>
<td>Value M.3</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>CAS N</td>
<td>NAME N</td>
<td>SMILES N</td>
<td>Value 1.N</td>
<td>Value 2.N</td>
<td>...</td>
<td>Value M.N</td>
</tr>
</tbody>
</table>
Import of database: Vertical import layout

• In this tutorial an example of vertical import of a database containing experimental information about Biochemical oxygen demand (BOD) and Bacterial reversed mutation assay (Ames) is shown.
• The excel file (Vertical import_ BOD and Ames.xlsx) is depicted below and it could be found in the following installation folder: C:\Program Files (x86)\Common Files\(Q)SAR Toolbox 4\Config\Examples

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CAS</td>
<td>NAME</td>
<td>Smiles</td>
<td>BOD</td>
<td>Ames</td>
</tr>
<tr>
<td>2</td>
<td>60-34-4</td>
<td>METHYLMETHANEDIONE</td>
<td>CN</td>
<td>0.2</td>
<td>Positive</td>
</tr>
<tr>
<td>3</td>
<td>50-29-3</td>
<td>DICHLORO_DIPHENYL_TREECHLORIDE</td>
<td>C(C)(N)(C)(C)(C)</td>
<td>7.1</td>
<td>Negative</td>
</tr>
<tr>
<td>4</td>
<td>50-32-8</td>
<td>BENZO PYRENE, 3,4'-BENZO</td>
<td>c12c3c4c5=Occcc3c1ccc5c</td>
<td>= 60.7</td>
<td>Negative</td>
</tr>
<tr>
<td>5</td>
<td>50-33-9</td>
<td>PHENYL BUTAZONE</td>
<td>C(=O)C(CCC)C(=O)NC=0</td>
<td>0.09</td>
<td>Negative</td>
</tr>
<tr>
<td>6</td>
<td>148-82-3</td>
<td>MELPHALAN</td>
<td>C=N=OcC1cc(C(N(C)C)CCCCC)</td>
<td>5</td>
<td>Positive</td>
</tr>
<tr>
<td>7</td>
<td>154-92-8</td>
<td>carbostyril</td>
<td>C=O(N[NCC(N)]N=O)NC=CC</td>
<td>0.09</td>
<td>Positive</td>
</tr>
<tr>
<td>8</td>
<td>6178-57-7</td>
<td>Benzo[c]azaridine</td>
<td>C(N)=O=O=cC3 OC3</td>
<td>1.63</td>
<td>Positive</td>
</tr>
<tr>
<td>9</td>
<td>62-75-9</td>
<td>NITROSODIMETHYLMINE</td>
<td>CN=N=O</td>
<td>80</td>
<td>Positive</td>
</tr>
<tr>
<td>10</td>
<td>91-59-8</td>
<td>2-NAPHTHYLAMINE</td>
<td>C12(c)(c)(n)(c)c=CC2</td>
<td>25</td>
<td>Positive</td>
</tr>
<tr>
<td>11</td>
<td>96-09-3</td>
<td>STYRENE OXIDE</td>
<td>c12C2OC2</td>
<td>cccc1</td>
<td>0.01</td>
</tr>
<tr>
<td>12</td>
<td>107-13-1</td>
<td>2-propenenitrile</td>
<td>c7c=O=cC</td>
<td>2</td>
<td>Positive</td>
</tr>
<tr>
<td>13</td>
<td>51-79-6</td>
<td>URETHANE</td>
<td>C(N)=O=OCC</td>
<td>1.8</td>
<td>Negative</td>
</tr>
<tr>
<td>14</td>
<td>53-96-3</td>
<td>2-ACETYLAMINOFLUORENE</td>
<td>c12=c3=cc(C)(N(C)=O)c3</td>
<td>cccc2</td>
<td>6.2</td>
</tr>
<tr>
<td>15</td>
<td>54-11-5</td>
<td>Pyridine, 3-(1-methyl-2-pyrrid)</td>
<td>c12CCC=NC3ccc1=CC</td>
<td>75</td>
<td>Positive</td>
</tr>
<tr>
<td>16</td>
<td>54-42-2</td>
<td>idoxuridine</td>
<td>C(=O)C(=O)C(C=0)C(C=0)C=0</td>
<td>0.09</td>
<td>Negative</td>
</tr>
<tr>
<td>17</td>
<td>55-38-9</td>
<td>FENTHION</td>
<td>c12SCC=OcOP=O=OCC=CC1</td>
<td>1.8</td>
<td>Negative</td>
</tr>
<tr>
<td>18</td>
<td>55-48-1</td>
<td>atropine sulphate</td>
<td>c12=O(c)c1ccoCC=CC</td>
<td>4.5</td>
<td>Negative</td>
</tr>
</tbody>
</table>
Import of database: Vertical import procedure

1. Go to Data module(1);
2. Click on Import button (2);
3. Click on Open file (3);
4. Select the file (Vertical import_ BOD and Ames.xlsx)(4);
5. Click on Open (5).
Import of database: Vertical import procedure

1. A preview of the imported file is shown (1);
2. The title of the imported file is also included;
3. Click on Next (3);
1. Select *Vertical* radio button (1).
2. Checkbox indicating the file for import includes Header row (2)
3. First three columns includes chemical ID information. These columns are automatically recognized by the system that they includes chemical ID. Recognition is displayed via assigning the correct ID identifiers on the second row (first column is recognized that it includes CAS#, second includes names and third includes SMILES). In this situation no action is needed from the user here.
4. However the columns with endpoint data are not recognized by system. In other words in the second row it is written “No endpoint selected”. In this respect in order to assign the endpoints to the respective column you need to click individually on each *No endpoint selected* button under the endpoints names (4)(see on next slide).
For **BOD**: Once click on **"Endpoint not selected"** a new window appears. First you need to select the hard path of the tree to which the data should be assigned. Open “Biodegradation node” (1)select “Ready Biodegradability” (2) from the endpoint tree and then to press Next (3).
For BOD:
Additionally select the Family/Scale from the drop-down menu (Biodegradability (%) (4)) and after that select the unit (% Biodegradability (%)) (5). Press Next (6).
For BOD: In order to define the endpoint more specifically please select BOD endpoint (7) from the drop-down menu. Then from the drop-down menu associated with test guideline select OECD301C (8).
Import of database: Vertical import procedure

For BOD: Once all the data fields are filled, press Finish (9).
For **BOD**: The edited fields are displayed in the main table (10).
**For Ames:**
1. Click on *No endpoint selected* (1);
2. Expand the Human Health hazard tree and select “Genetic toxicity” (2).
3. Click on *Next* (3)
For **Ames**: From the drop-down menu of the *Endpoint* field (4) select Gene mutation (5).
For **Ames**: From the drop-down menu of **Type of method**, select In vitro (6) and from **Test type** field select: Bacterial reverse Mutation assays (e.g. Ames test) (7).
For **Ames**:
Select **Test organism (species)**: Salmonella typhimurium (8) and for **Metabolic activation** select Without S9 (9).
For **Ames**: Select Strain: TA 100 (10) for instance and then click Finish (11).
For Ames: The edited fields are displayed in the main table (12). Finally, click on Import (13).
Import of database: Vertical import procedure

- The import process could take a couple of minutes;
- An informative message is displayed when it is completed;
- Click on OK (1).
The new database is displayed in the Databases panel (1) in two sections: *Environmental fate and transport* group and *Human health hazard* group (2);

Right-click menu (3) is implemented where you can see the database statistics or delete the database.
Outlook

• Aim
• Keywords
• Definition of Database and Inventory
• Import:
  • Import of database
    • Vertical import
    • Horizontal import
  • Import of inventory
  • Import via IUCLID
• Export Data matrix
Import of database: Horizontal import layout

- In this tutorial an example of horizontal import of a database containing experimental information related to genotoxicity is shown.
- The excel file (Horizontal import_Genotoxicity.xlsx) is depicted below and could be found in the following installation folder: C:\Program Files (x86)\Common Files\(Q)SAR Toolbox 4\Config\Examples
1. Go to **Data** module (1);
2. Click on **Import** (2);
3. Type in the name of the database as you would want it to be displayed in Toolbox (3). Otherwise, the name of the file will be used as the name of the database.
4. Click on **Open file** (4);
5. Select the file *(Horizontal import_genotoxicity.xls)* (5);
6. Click on **Open** (6).
1. A preview of the imported file is shown (1);
2. Click on Next (2);
Select **Horizontal** radio button (1)
- Column with chemical ID is automatically recognized by the system (for more details see slide 15)
- When there are fields from the original file which cannot be mapped to the labels existing in Toolbox, burgundy colored messages are displayed on the top (2) and Undefined is written below the data which is not mapped (3). In order to map correctly the fields see next slide.
1. Open the drop-down menu (1), type in “end” in the filter (2), select EndpointPath (3).

2. The selected label has to correspond to the one in the original file and it is also written in the top message (4).
Import of database: Horizontal import procedure

1. Open the drop-down menu (1), type in “data” in the filter (2), select Data.MeanValue (3).
2. The selected label has to correspond to the one in the original file and it is also written in the top message (4).
Once all fields are mapped, the burgundy colored messages on top disappear.

You can use the scrollbar (1) to check all columns, their titles and content.

Click on Import (2)
The import process could take a couple of minutes;
• An informative message is displayed when it is completed;
• Click on **OK** (1).
View of imported database

• The new database is displayed in the Databases panel (1) in the **Human health hazard** group (2).
• Right-click menu (3) is implemented where you can see the database statistics or delete the database.
• The software automatically adds a numeration in the name ("1" in this example) in case the same database is imported for the second time.
Outlook

• Aim
• Keywords
• Definition of Database and Inventory
• **Import of database:**
  - Vertical import
  - Horizontal import
  - **Supporting information**
• Import of inventory
• Export Data matrix
Supporting information

• A table of the most important endpoints implemented in Toolbox could be found in *F1 Help: D.3.4.4. Supporting Information*.

• The information in the table aims to facilitate the users’ work when importing new databases.
Outlook

• Aim
• Keywords
• Definition of Database and Inventory
• **Import:**
  • Import of database
    • Vertical import
    • Horizontal import
  • **Import of inventory**
  • Import via IUCLID
• Export Data matrix
Import of inventory

Layout

• As already presented an inventory is a collection of chemicals without any data.

• In the current example import of chemicals along with their external identifiers (custom ID) will be illustrated.
• Import layout contains two main sections: substance information (1) and column with external ID information (custom ID number)(2);
• The external ID number could be different types: strings only, integer (numbers) only or combination of integer and strings (3)
• The imported file can be in **xlsx** or **tsv** format
• An example file could be found in the following installation folder: C:\Program Files (x86)\Common Files\(Q)SAR Toolbox 4\Config\Examples
Import of inventory
Procedure

1. Go to Data module (1);
2. Click on Import (2);
3. Tick “Import as inventory” (3);
4. Type in the name of the inventory as you would want it to be displayed in Toolbox (4). Otherwise the name of the file will be used as the name of the inventory.
5. Click on Open file
6. Select the file (Import_Custom Inventory_Custom IDs.xlsx)(6),
7. Click Open (7) then a message appears informing that only the first worksheet will be used for import, click OK
1. A preview of the imported file is shown (1);
2. Click on *Next* (2);
1. On the second top row are displayed the chemical identifiers as they will be shown in Toolbox (1);

2. The last column includes external ID and it is not recognized by the system (the second row is named “Undefined”). Open the drop-down menu and type “custo” (2). Two options appeared:
   - Add as custom id(integer)
   - Add as custom id(string)

3. If your external IDs are only numbers (integer) you should select the first option. If your custom IDs are letters only you should select the second option. In case your IDs are combination of letters and numbers, you should select again the second option. In this case external IDs are combination of letters and strings, so second option is selected (3).

4. Click **Import** (4);
The import process could take a couple of minutes;
• An informative message is displayed when it is completed;
• Click on OK (1).
The new inventory (1) is displayed in the Inventories panel (2).
Right-click menu (3) is implemented where you can or delete the inventory.
The software automatically adds a numeration in the name ("1" in this example) in case the same database is imported for the second time.
Outlook

- **Aim**
- **Keywords**
- **Definition of Database and Inventory**
- **Import:**
  - Import of database
    - Vertical import
    - Horizontal import
  - Import of inventory
  - **Import of data from IUCLID to TB**
  - Export Data matrix
Import of data from IUCLID to Toolbox

1. Go to **Data** module (1);
2. Click on **IUCLID6** button (2);
3. Select New database (3);
4. Write the database name (4);
5. Click on **Next** (5).
Import of data from IUCLID to Toolbox

1. Write **IUCLID Server name** (1);
2. Next Port(2);
3. Next Username (3);
4. Password: root(4);
5. Click on **Next** (5).

**Note:**
In case you don't know your IUCLID account, please ask your administrator.
Import of data from IUCLID to Toolbox

1. Get All Substances (1).
2. Select the substance which you would like to import in the Toolbox (2)
3. Finish (3)
Import of data from IUCLID to Toolbox

1. Import completed (1).
1. Click on **Input** module, then click on **Database** button (1).

2. Select the imported IUCLID database (in our case named IUCLID file) in order to load it in the data matrix (2) and click **OK** (3).
Outlook

• Aim
• Keywords
• Definition of Database and Inventory
• Import data:
  • Import of database
    • Vertical import
    • Horizontal import
  • Import of inventory
  • Import IUCLID
• Export data
Export
Types of export

• Data could be exported via:
  • Data matrix export
  • IUCLID Export*

*IUCLID Export is allowed for predictions only. An example of export is shown in the tutorial “Predicting aquatic toxicity to daphnia by trend analysis”
Data matrix export:

Allows to export chemicals with data and supporting information (e.g. profiling results, 2D/3D parameters, molecular formula, etc.) available on the Data matrix in text format organized in a horizontal layout.

Two options to export:

- Export data from a row
- Export whole data matrix

In this tutorial only the export of the **whole data matrix** is shown as it encompasses the first option as well.
1. Go to **Input** module (1);
2. Go to Select (2)
3. Select from Database (3);
4. Select *Aquatic OASIS*(4);
5. Click on **OK**(5).
1. Select chemicals from the first row by holding Ctrl button and click over the chemicals (1);
2. Click on OK(2).
Data matrix Export: Collecting data

The following steps need to be executed to collect the experimental data for the chemicals available on data matrix:

1. Go to **Data** module (1);
2. Select **Aquatic OASIS** (2);
3. Click on **Gather** (3);
4. Click on **OK** in the Read Data window (4).
5. Click on **OK** in the information window (5).
1. Go to **Profiling** module (1)
2. Select **Acute aquatic toxicity MOA by OASIS** profiler and click **Apply** (2)
3. Perform right click over the level of Ecotoxicological Information and select **Export data matrix** (3)
1. Keep the Aquatic toxicity level selected (1); Unselect Sediment and Terrestrial toxicity (not shown).
2. Expand the Profile level (2) and then select Acute aquatic toxicity MOA by OASIS;
3. Click on Export (3).
Data matrix Export

1. Specify the name of the file (e.g. Export data) (1). The file is saved in csv format.
2. Click on Save (2).
3. Click OK in the information message (3).
The file can be opened in excel. It contains the following main sections: **chemical identity of each chemical** (red, 1), **endpoint** (pink, 2) **experimental data and unit** (purple, 3), metadata (blue, 4), profiler results (green, 5). Here, if the chemical has several experimental data each one is listed on a separate row (e.g. CAS 98-01-1).
Congratulations!

• You have now been introduced to the functionality of import and export of data;
• You have now been introduced to the import of data from IUCLID to the Toolbox;
• Note, proficiency comes with practice!