QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

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.

CAS #	NAME	SMILES	Experiment 1	Experiment 2	...	Experiment M
CAS 1	NAME 1	SMILES 1	Value 1.1	Value 2.1	...	Value M.1
CAS 2	NAME 2	SMILES 2	Value 1.2	Value 2.2	...	Value M.2
CAS 3	NAME 3	SMILES 3	Value 1.3	Value 2.3	...	Value M.3
...	...	...	...	...	...	...
CAS N	NAME N	SMILES N	Value 1.N	Value 2.N	...	Value M.N

1

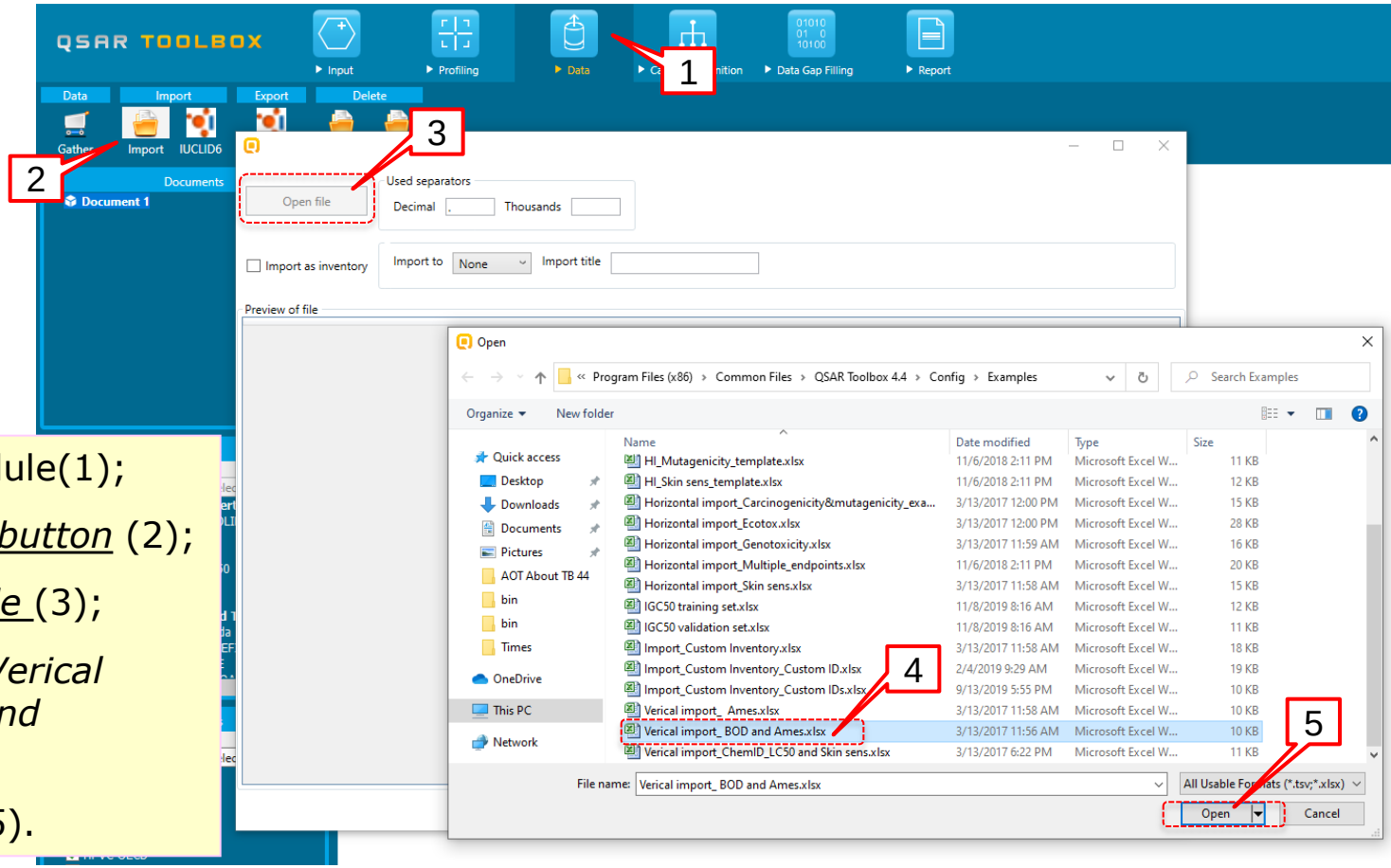
2

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*

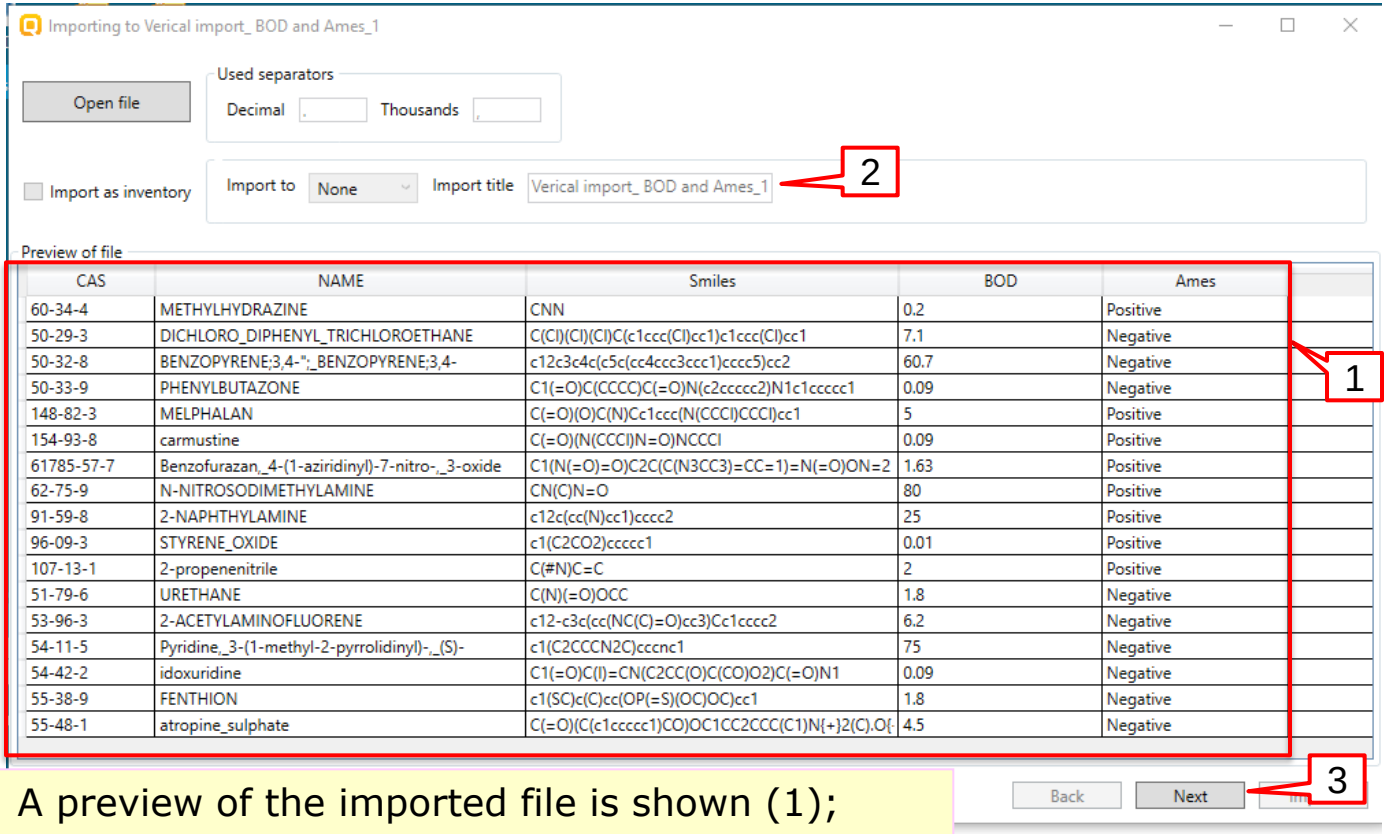
	A	B	C	D	E
1	CAS	NAME	Smiles	BOD	Ames
2	60-34-4	METHYLHYDRAZINE	CNN	0.2	Positive
3	50-29-3	DICHLORO_DIPHENYL_TRICHL	C(Cl)(Cl)(Cl)C(c1ccc(Cl)cc1)c1ccc(	7.1	Negative
4	50-32-8	BENZOPYRENE;3,4-";_BENZOI	c12c3c4c(c5c(cc4ccc3ccc1)cccc5)c	60.7	Negative
5	50-33-9	PHENYLBUTAZONE	C1(=O)C(CCCC)C(=O)N(c2ccccc2)	0.09	Negative
6	148-82-3	MELPHALAN	C(=O)(O)C(N)Cc1ccc(N(CCCl)CCC	5	Positive
7	154-93-8	carmustine	C(=O)(N(CCCl)N=O)NCCCl	0.09	Positive
8	61785-57-7	Benzofurazan,_4-(1-aziridiny	C1(N(=O)=O)C2C(C(N3CC3)=CC=1	1.63	Positive
9	62-75-9	N-NITROSODIMETHYLAMINE	CN(C)N=O	80	Positive
10	91-59-8	2-NAPHTHYLAMINE	c12c(cc(N)cc1)cccc2	25	Positive
11	96-09-3	STYRENE_OXIDE	c1(C2CO2)ccccc1	0.01	Positive
12	107-13-1	2-propenenitrile	C(#N)C=C	2	Positive
13	51-79-6	URETHANE	C(N)(=O)OCC	1.8	Negative
14	53-96-3	2-ACETYLAMINOFLUORENE	c12-c3c(cc(NC(C)=O)cc3)Cc1cccc2	6.2	Negative
15	54-11-5	Pyridine,_3-(1-methyl-2-pyrr	c1(C2CCCN2C)cccnc1	75	Negative
16	54-42-2	idoxuridine	C1(=O)C(I)=CN(C2CC(O)C(CO)O2	0.09	Negative
17	55-38-9	FENTHION	c1(SC)c(C)cc(OP(=S)(OC)OC)cc1	1.8	Negative
18	55-48-1	atropine_sulphate	C(=O)(C(c1ccccc1)CO)OC1CC2CC(	4.5	Negative
19					

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 (*Verical import_ BOD and Ames.xlsx*)(4);
5. Click on *Open* (5).

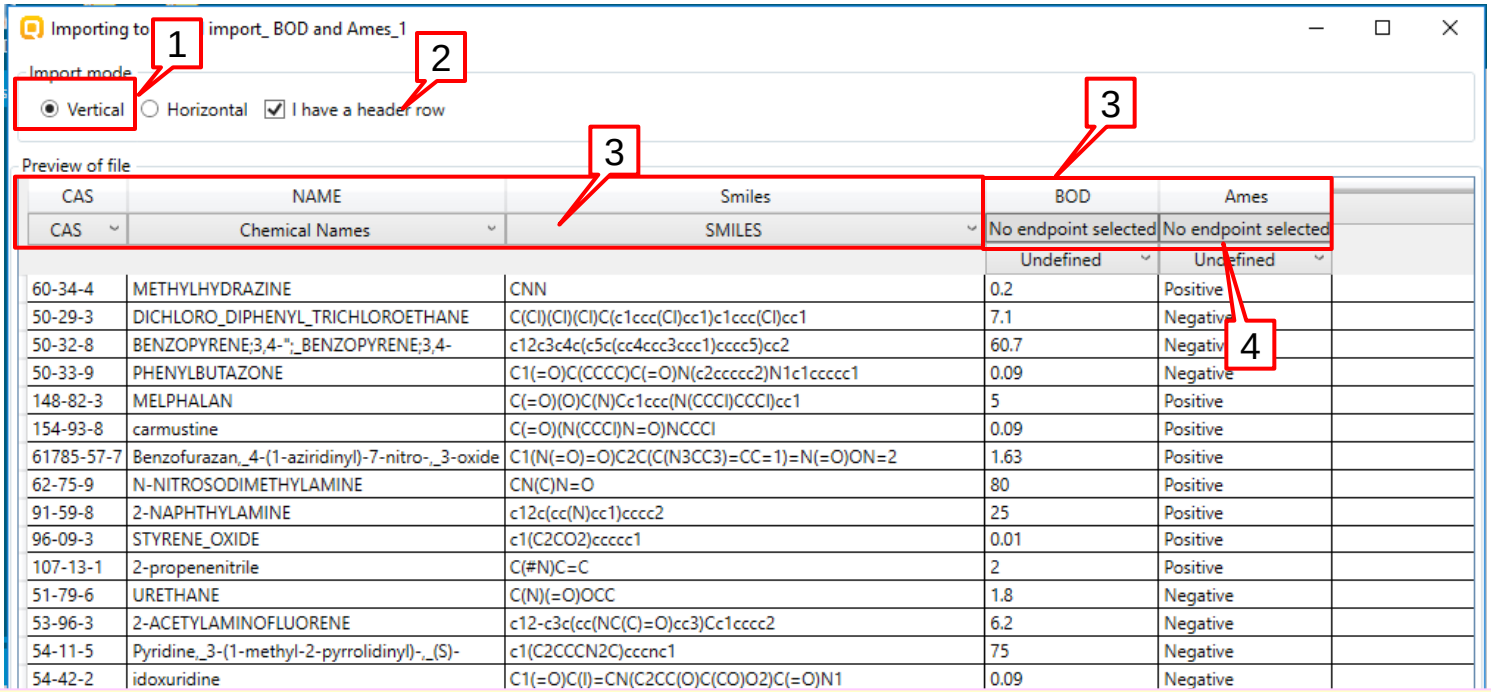
Import of database: Vertical import procedure



CAS	NAME	Smiles	BOD	Ames
60-34-4	METHYLHYDRAZINE	CNN	0.2	Positive
50-29-3	DICHLORO_DIPHENYL_TRICHLOROETHANE	C(Cl)(Cl)(Cl)C(c1ccc(Cl)cc1)c1ccc(Cl)cc1	7.1	Negative
50-32-8	BENZOPYRENE;3,4-",_BENZOPYRENE;3,4-	c12c3c4c(c5c(cc4ccc3ccc1)cccc5)cc2	60.7	Negative
50-33-9	PHENYLBUTAZONE	C1(=O)C(CCCC)C(=O)N(c2ccccc2)N1c1ccccc1	0.09	Negative
148-82-3	MELPHALAN	C(=O)(O)C(N)Cc1ccc(N(CCCl)CCCl)cc1	5	Positive
154-93-8	carmustine	C(=O)(N(CCCl)N=O)NCCCl	0.09	Positive
61785-57-7	Benzofurazan,_4-(1-aziridinyl)-7-nitro-,_3-oxide	C1(N(=O)=O)C2C(C(N3CC3)=CC=1)=N(=O)ON=2	1.63	Positive
62-75-9	N-NITROSODIMETHYLAMINE	CN(C)N=O	80	Positive
91-59-8	2-NAPHTHYLAMINE	c12c(cc(N)cc1)cccc2	25	Positive
96-09-3	STYRENE_OXIDE	c1(C2CO2)ccccc1	0.01	Positive
107-13-1	2-propenenitrile	C(#N)C=C	2	Positive
51-79-6	URETHANE	C(N)(=O)OCC	1.8	Negative
53-96-3	2-ACETYLAMINOFLUORENE	c12-c3c(cc(NC(C)=O)cc3)Cc1cccc2	6.2	Negative
54-11-5	Pyridine,_3-(1-methyl-2-pyrrolidinyl)-,_(S)-	c1(C2CCCN2C)cccnc1	75	Negative
54-42-2	idoxuridine	C1(=O)C(I)=CN(C2CC(O)C(CO)O2)C(=O)N1	0.09	Negative
55-38-9	FENTHION	c1(SC)c(C)cc(OP(=S)(OC)OC)cc1	1.8	Negative
55-48-1	atropine_sulphate	C(=O)(C(c1ccccc1)CO)OC1CC2CCC(C1)N{+}2(C).O{-	4.5	Negative

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);

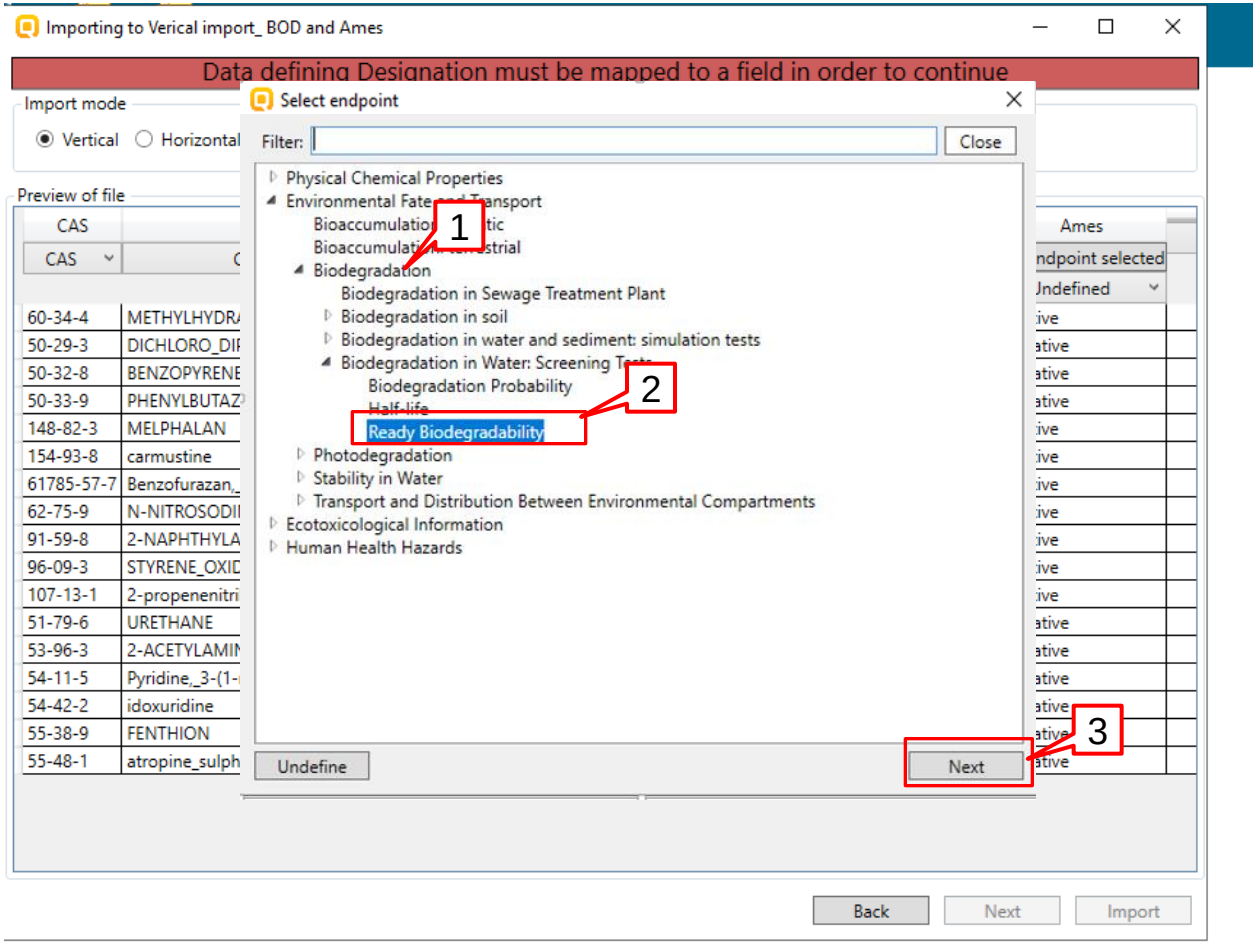
Import of database: Vertical import procedure



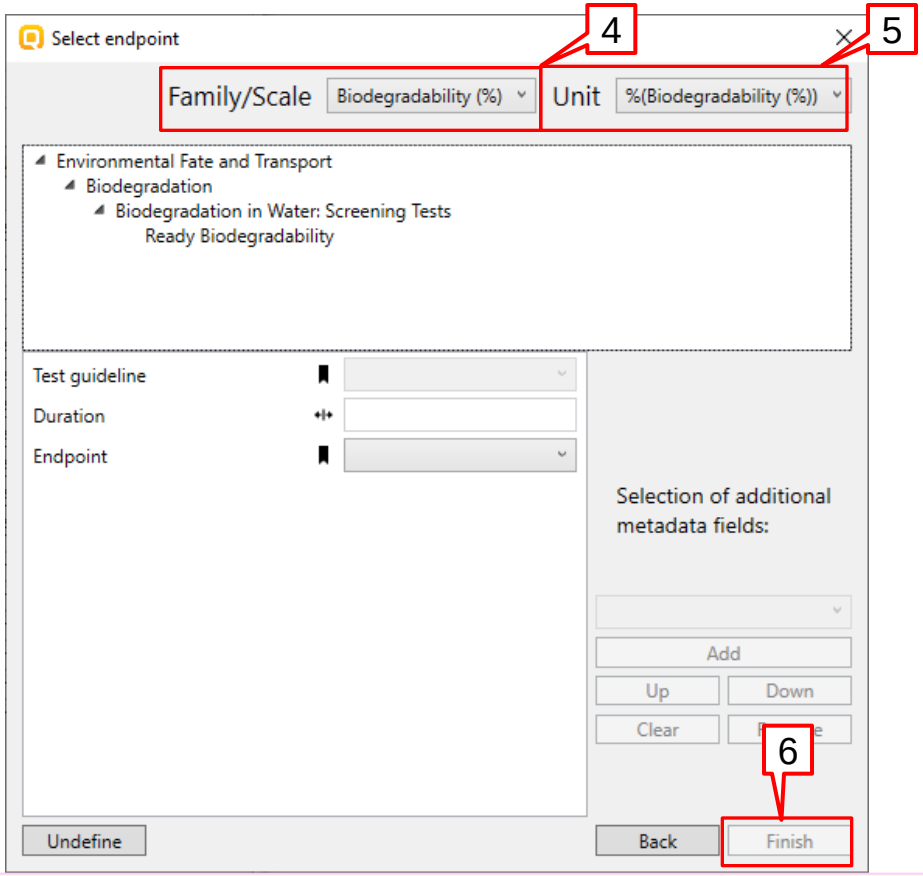
1. Select *Vertical* radio button (1).
2. Checkbox indicating the file for import includes Header row (2)
3. First three columns includes chemical ID information. These columns are automatically recognized by the system that they includes chemical ID. Recognition is displayed via assigning the correct 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).

Import of database: Vertical import procedure

*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).

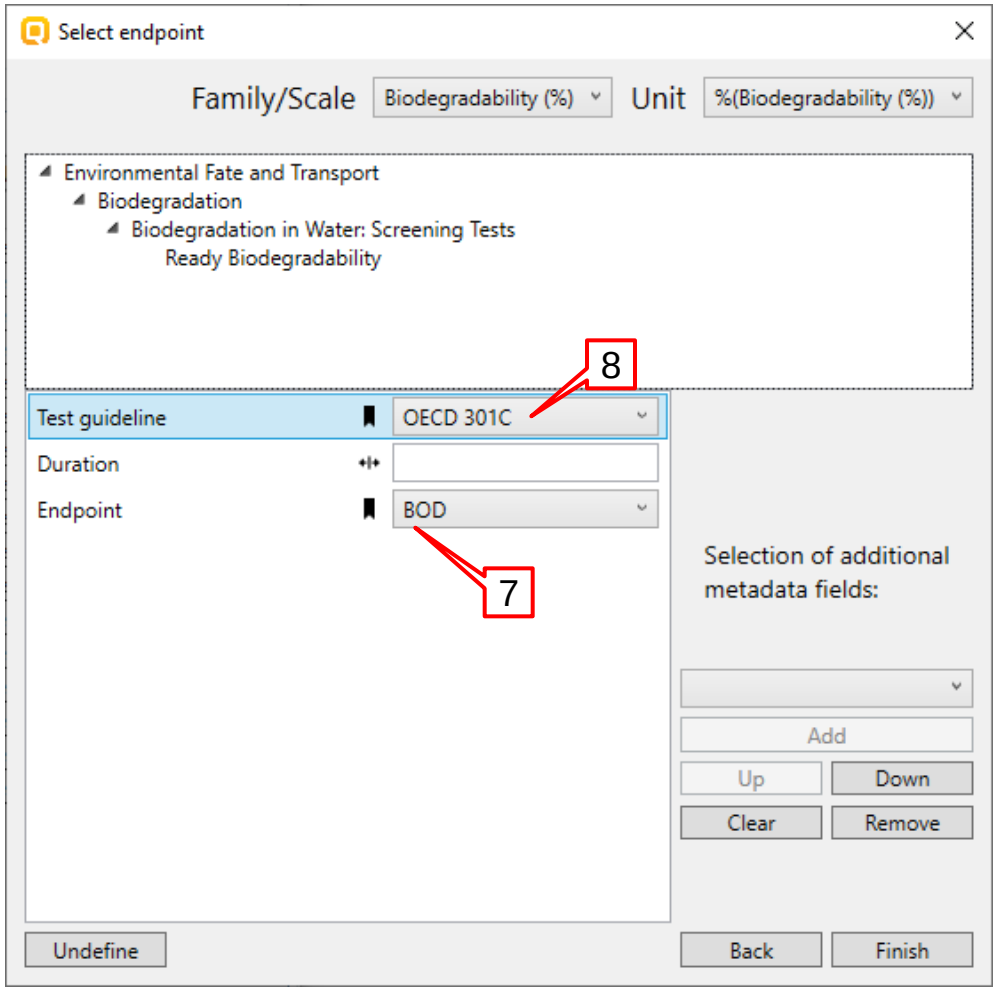


Import of database: Vertical import procedure



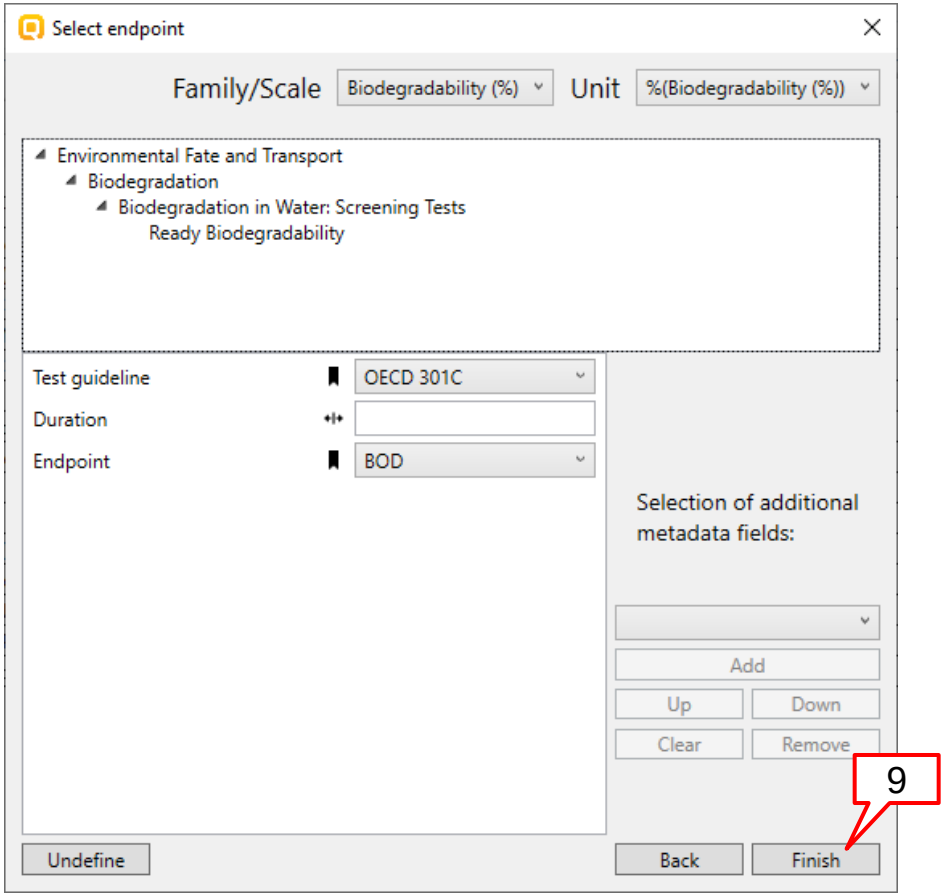
*For **BOD**:*
Additionally select the Family/Scale from the drop-down menu (Biodegradability (%)(4)) and after that select the unit (% Biodegradability(%))(5). Press *Next* (6).

Import of database: Vertical import procedure



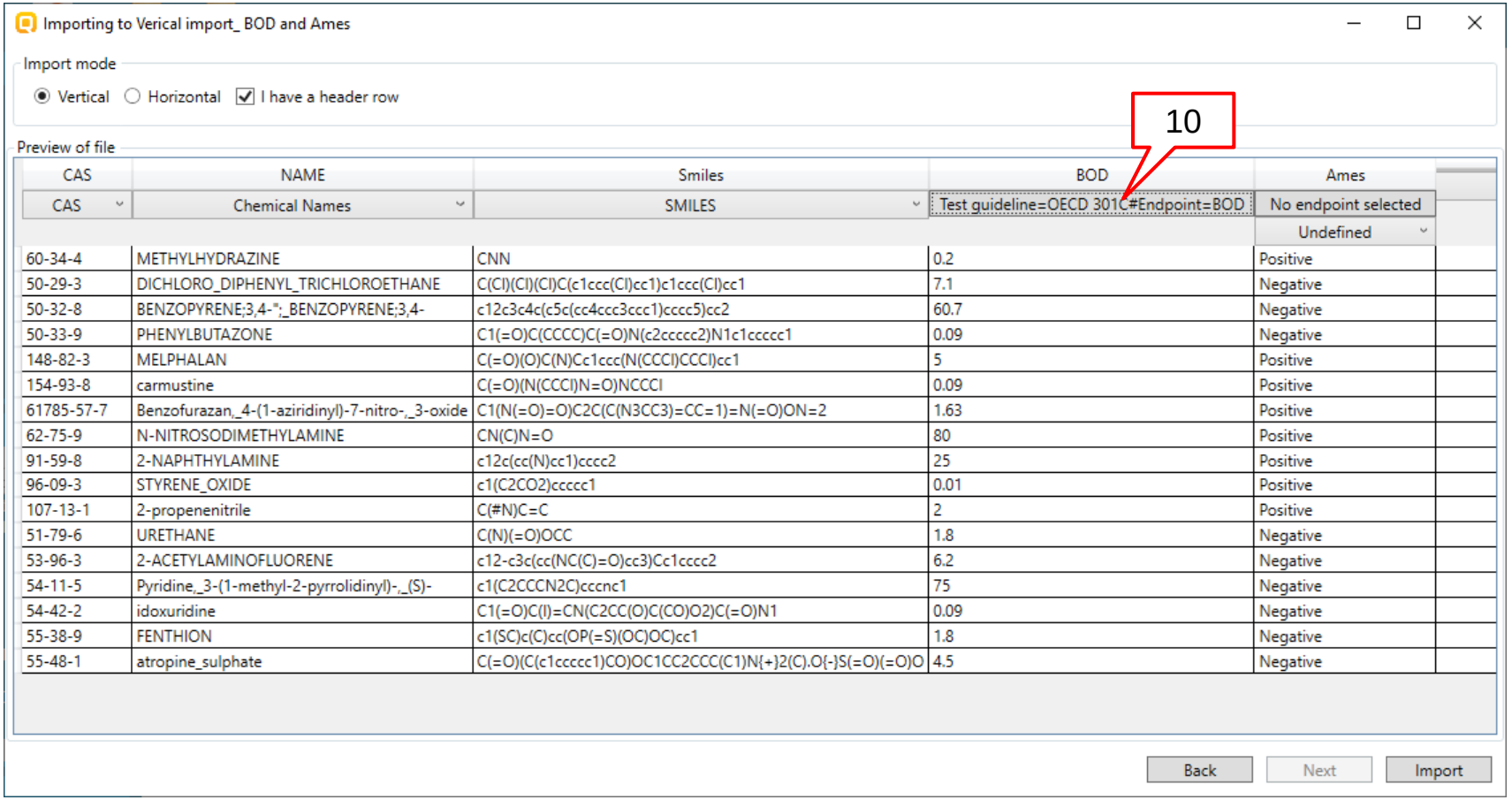
*For **BOD***: In order to define the endpoint more specifically please select *BOD endpoint* (7) from the drop-down menu. Then from the drop-down menu associated with test guideline select *OECD301C* (8).

Import of database: Vertical import procedure



*For **BOD***: Once all the data fields are filled , press *Finish* (9).

Import of database: Vertical import procedure

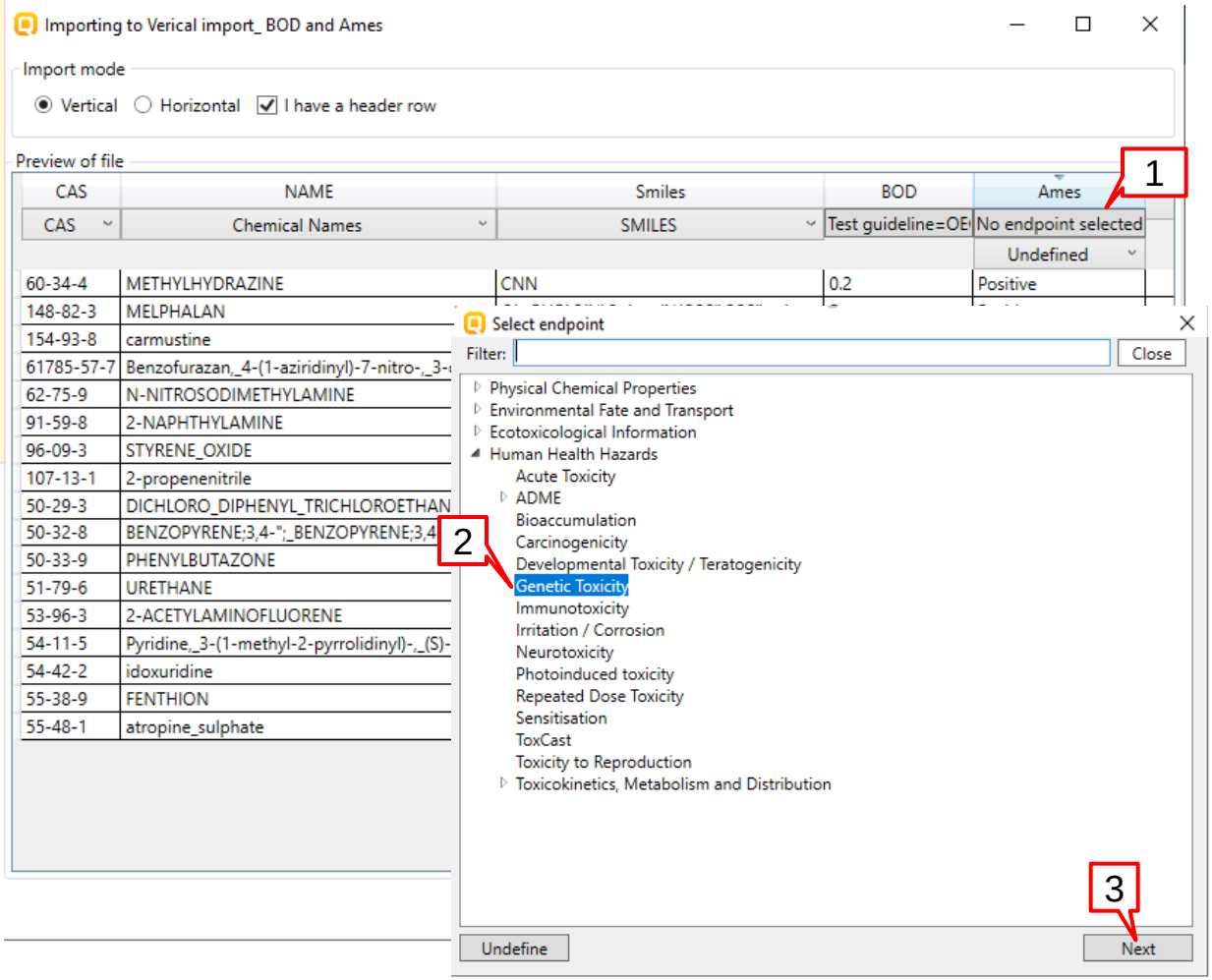


*For **BOD***: The edited fields are displayed in the main table (10).

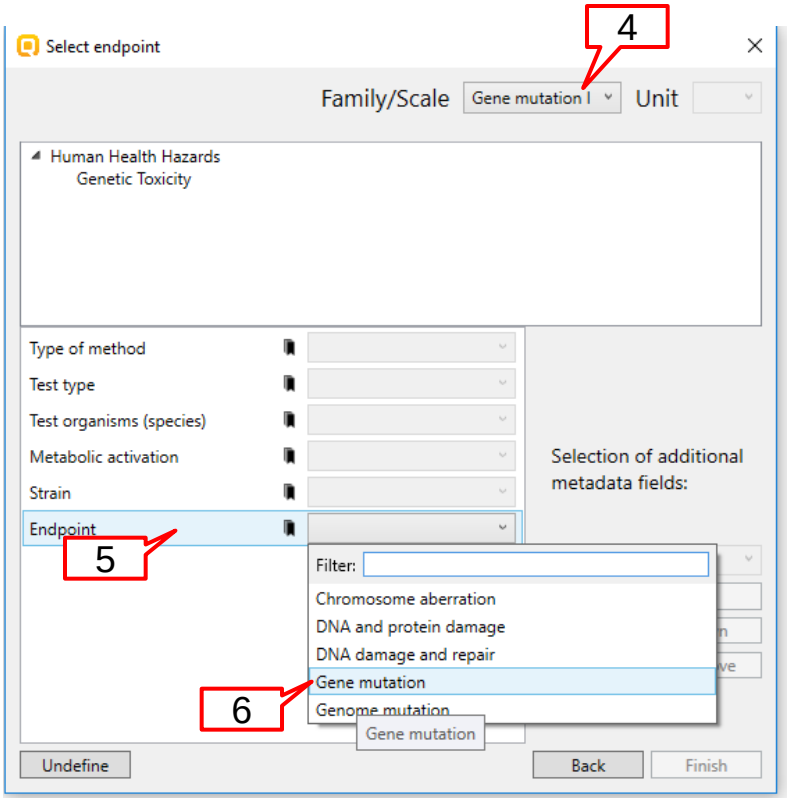
Import of database: Vertical import procedure

*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)

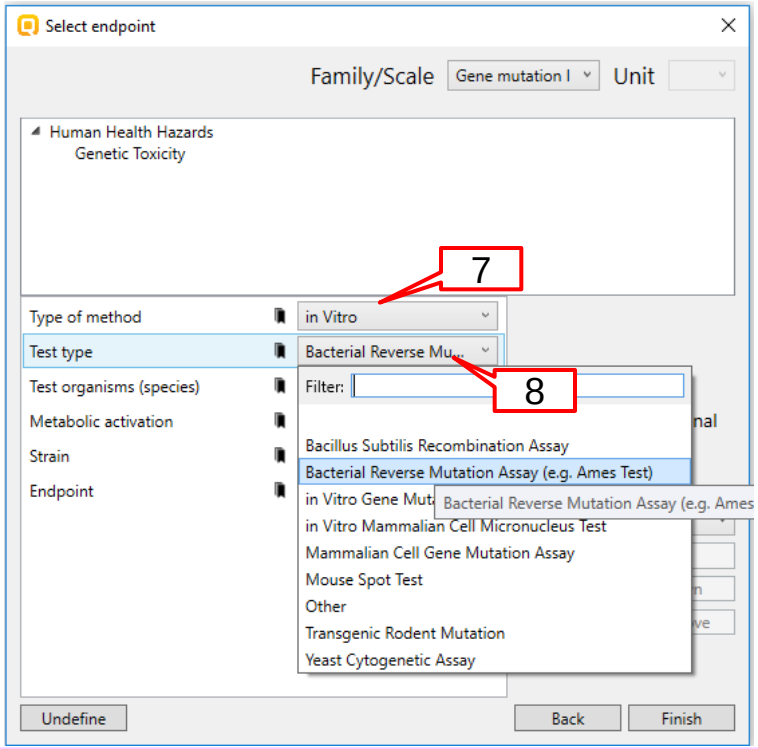


Import of database: Vertical import procedure



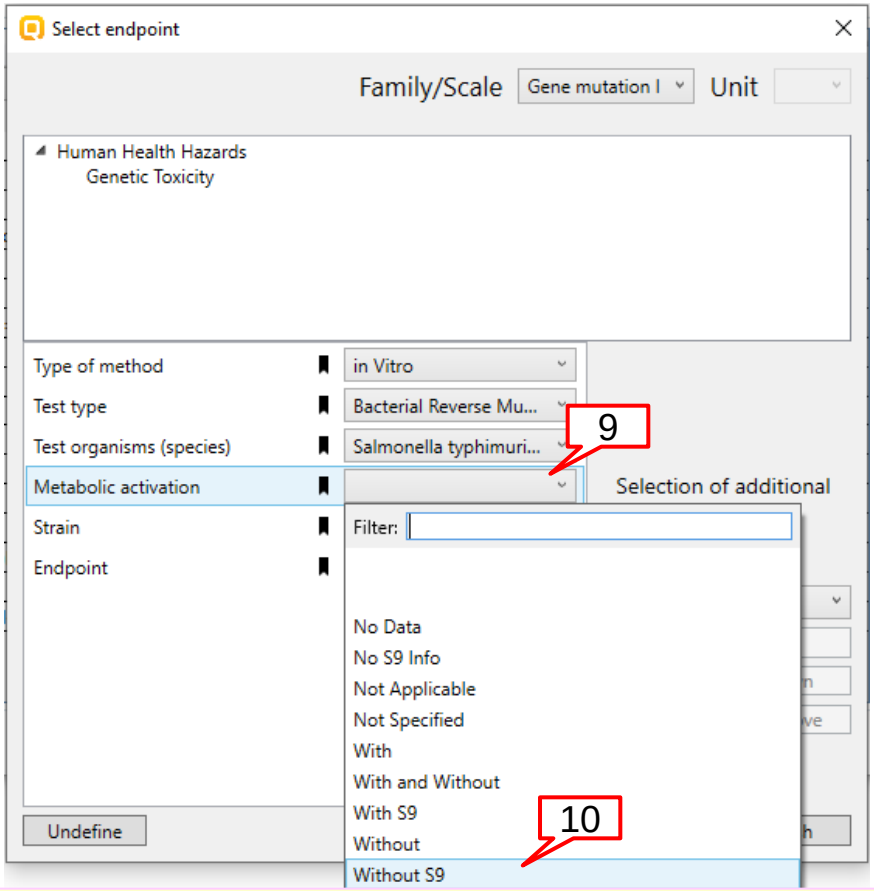
*For **Ames***:
Select scale Gene mutation I (4). From the drop-down menu of the *Endpoint* field (5) select Gene mutation(6).

Import of database: Vertical import procedure



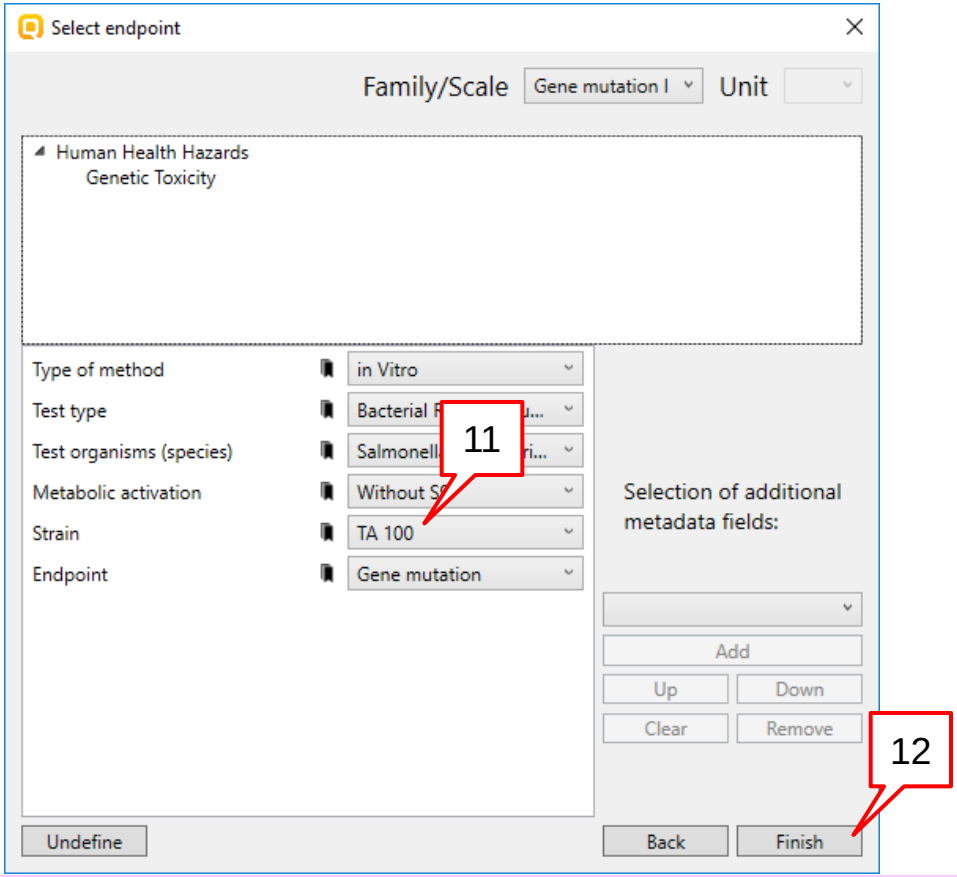
*For **Ames***:
From the drop-down menu of *Type of method*, select In vitro (7) and from *Test type* field select: Bacterial reverse Mutation assays (e.g. Ames test) (8).

Import of database: Vertical import procedure



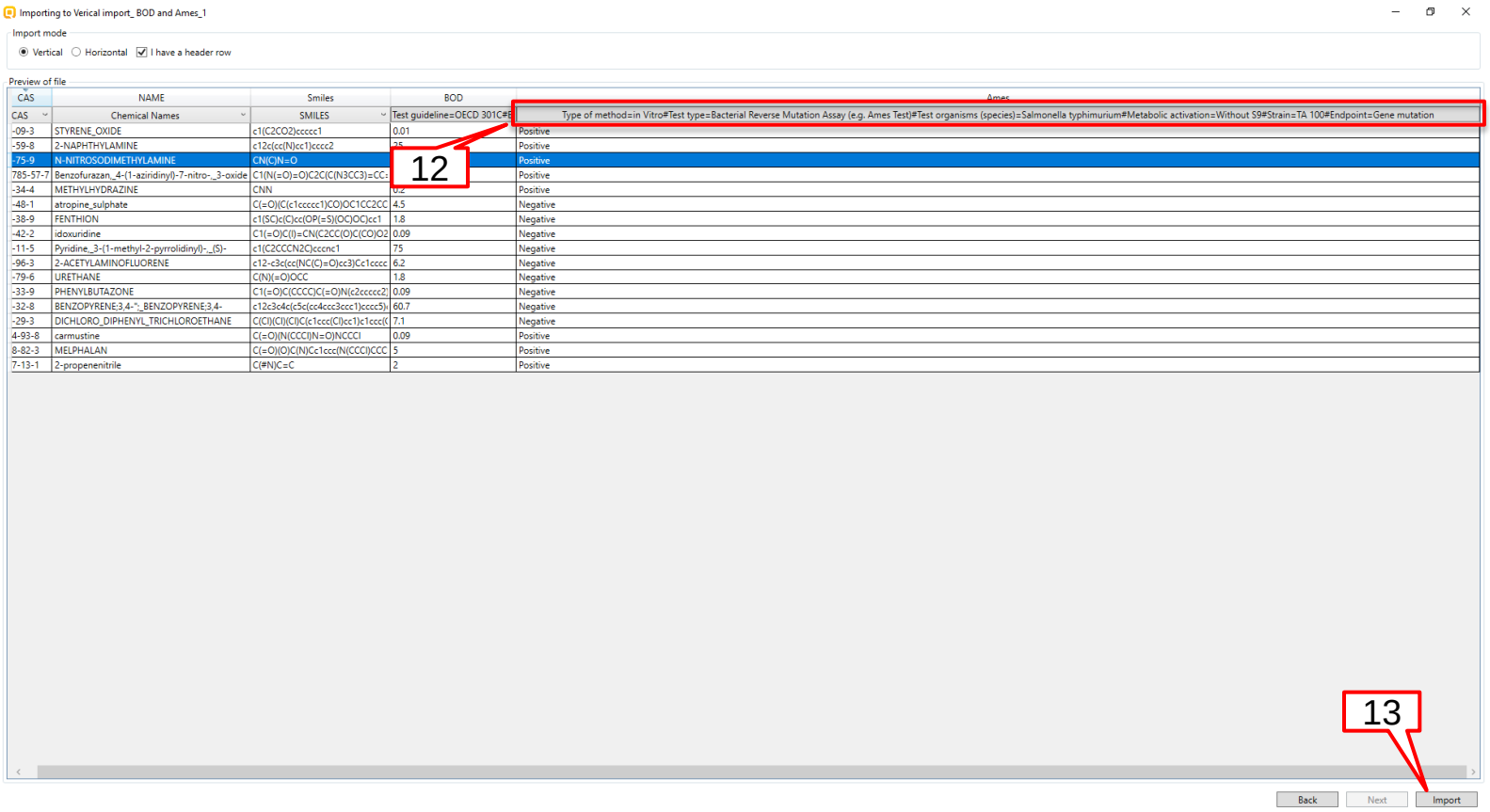
*For **Ames***:
Select *Test organism (species)*: Salmonella typhimurium (9) and for *Metabolic activation* select Without S9 (10).

Import of database: Vertical import procedure



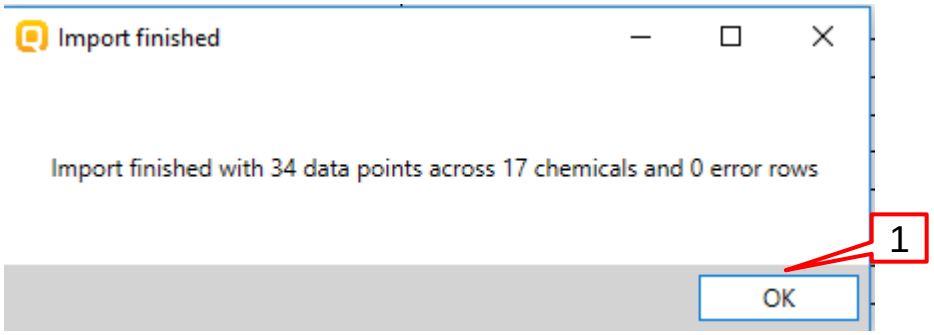
*For **Ames***:
Select Strain: TA 100 (11) for instance and then click Finish (12).

Import of database: Vertical import procedure



*For **Ames***: The edited fields are displayed in the main table (12).
Finally, click on *Import* (13).

Import of database: Vertical import procedure

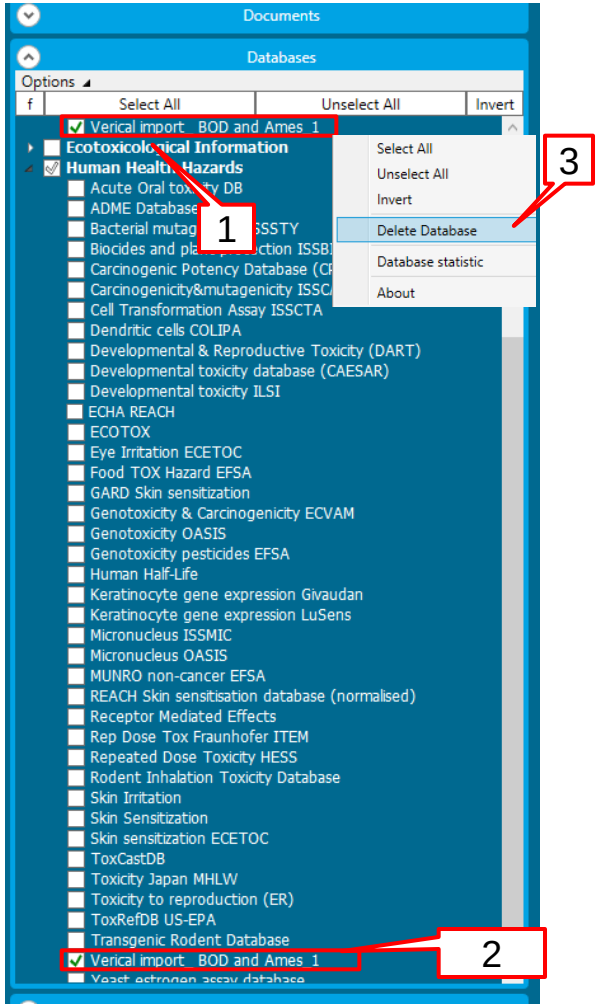


- The import process could take a couple of minutes;
- An informative message is displayed when it is completed;
- Click on *OK* (1).

View of imported database

- 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

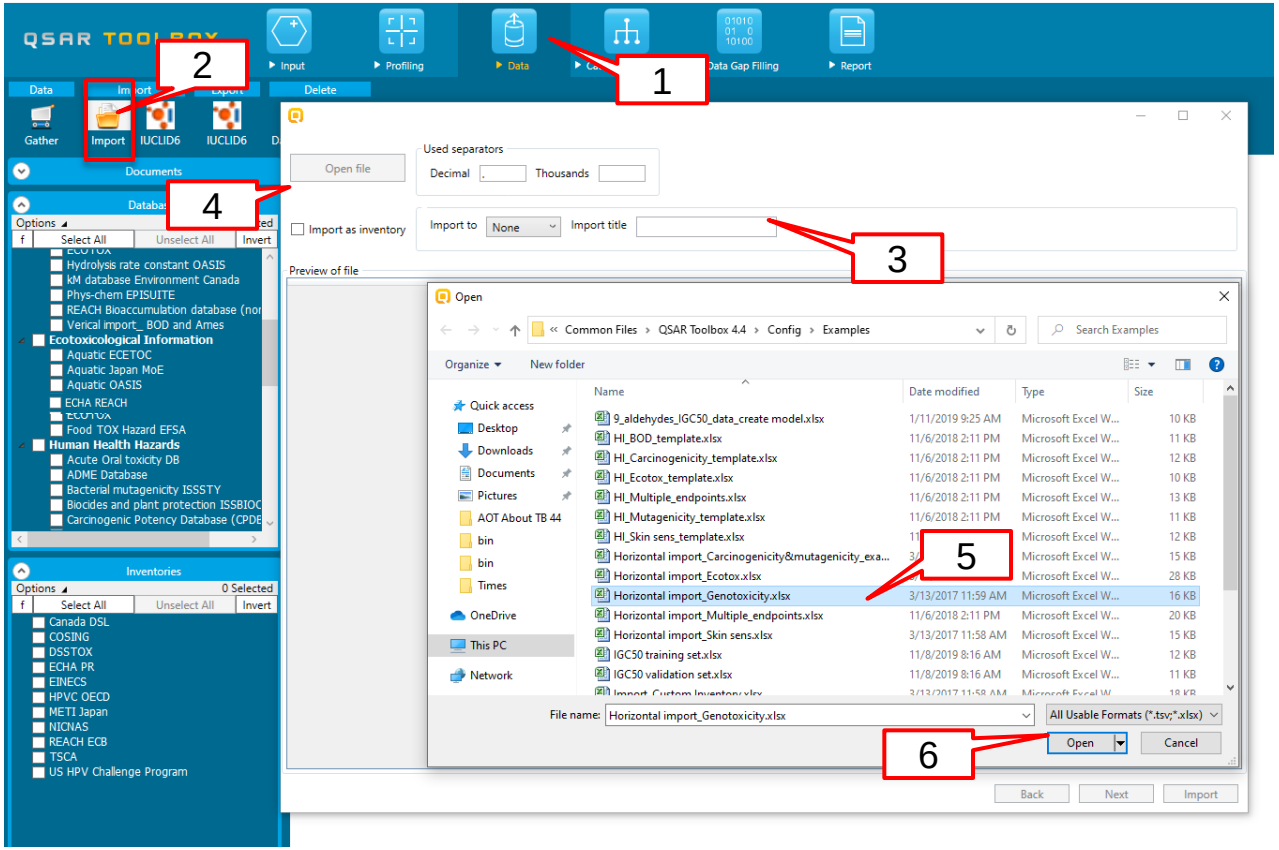
- In this tutorial an example of horizontal import of a database containing experimental information related to genotoxicity is shown.
- The excel file (Horizontal import_Genotoxicity.xlsx) is depicted below and it could be found in the following installation folder: C:\Program Files (x86)\Common Files\(Q)SAR Toolbox 4\Config\Examples

F1: Data Mean value/Scale value

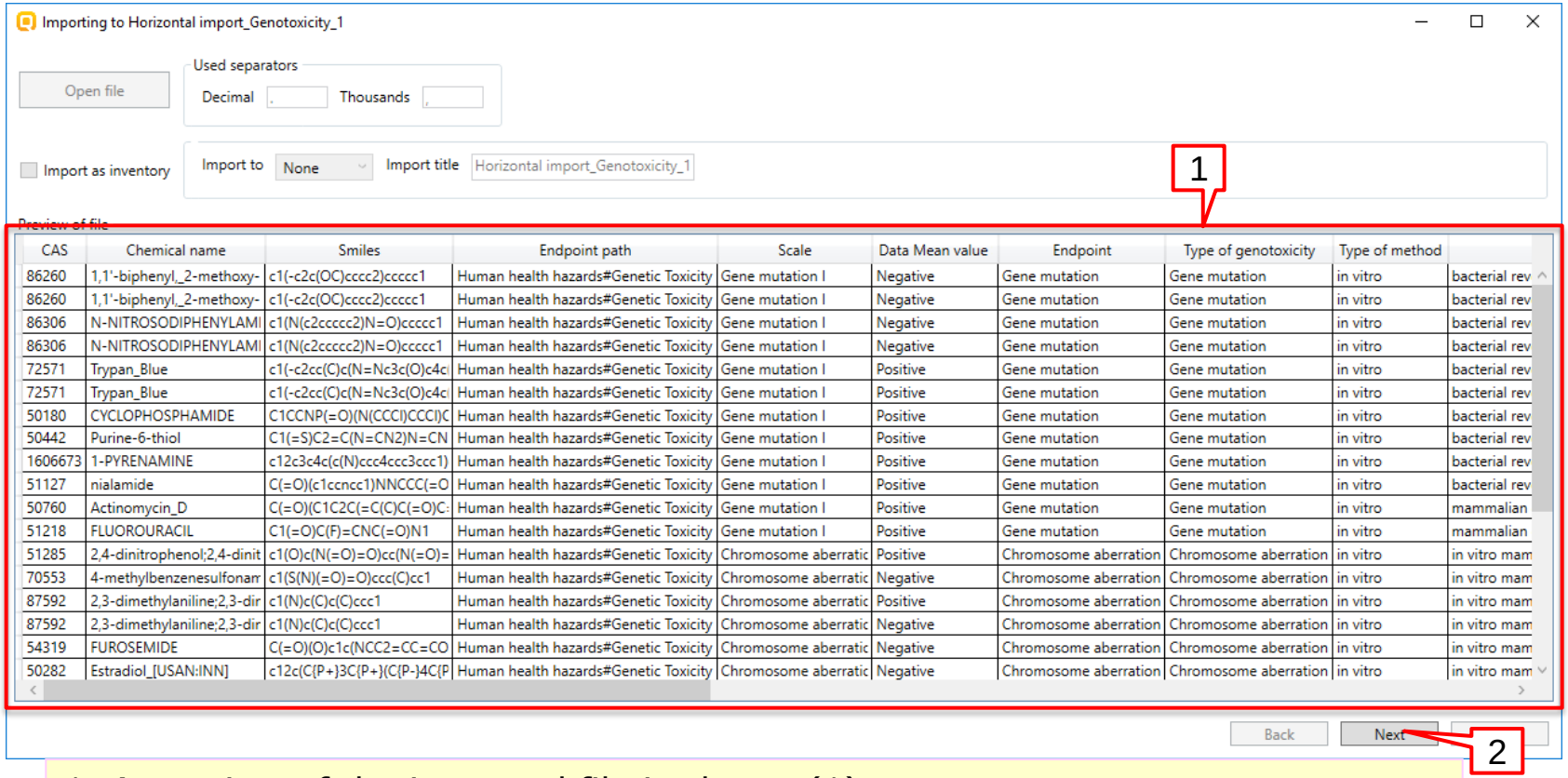
	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Data Mean value/Scale value	Endpoint	Type of genotoxicity	Type of method	Test type	Metabolic activation	Test organisms (species)	Strain	Source of methabolic system	Author	Title	Year	Chemical info
2	Negative	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	with S9	Salmonella typhimurium	TA 100	rat	U.S. Environ	The Salmonella typhimuri	1996	US_GTox
3	Negative	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	without S9	Salmonella typhimurium	TA 100		U.S. Environ	The Salmonella typhimuri	1996	US_GTox
4	Negative	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	with S9	Salmonella typhimurium	TA 100	rat	U.S. Environ	The Salmonella typhimuri	1996	US_GTox
5	Negative	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	without S9	Salmonella typhimurium	TA 100		U.S. Environ	The Salmonella typhimuri	1996	US_GTox
6	Positive	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	with S9	Salmonella typhimurium	TA 98	hamster	National Car	Short-Term Testing Progra	2000	NCI-STTP
7	Positive	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	without S9	Salmonella typhimurium	TA 98		National Car	Short-Term Testing Progra	2000	NCI-STTP
8	Positive	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	with S9	Salmonella typhimurium	TA 1535	mouse	U.S. Environ	The Salmonella typhimuri	1996	US_GTox
9	Positive	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	without S9	Salmonella typhimurium	TA 1535		U.S. Environ	The Salmonella typhimuri	1996	US_GTox
10	Positive	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	with S9	Salmonella typhimurium	TA 1537	rat	U.S. Environ	The Salmonella typhimuri	1996	US_GTox
11	Positive	Gene mutation	Gene mutation	in vitro	bacterial reverse muta	without S9	Salmonella typhimurium	TA 1535		U.S. Environ	The Salmonella typhimuri	1996	US_GTox
12	Positive	Gene mutation	Gene mutation	in vitro	alian cell gene mutation assay		Mouse Lymphoma cells			Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
13	Positive	Gene mutation	Gene mutation	in vitro	alian cell gene mutation assay		Mouse Lymphoma cells			Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
14	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chromosome aberra		Chinese Hamster	chinese hams		Japan Minist	Japan Ministry of Health L		
15	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chromosome aberra		Chinese Hamster	chinese hams		Japan Minist	Japan Ministry of Health L		
16	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chromosome aberra		Chinese Hamster	chinese hams		Japan Minist	Japan Ministry of Health L		
17	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chromosome aberra		Chinese Hamster	chinese hams		Japan Minist	Japan Ministry of Health L		
18	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	with S9			rat	LSIC	Japan-Danish EPA Inventc	2000	
19	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	without S9				Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
20	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	with S9			rat	LSIC	Japan-Danish EPA Inventc	2000	
21	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	without S9				Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
22	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	with S9			rat	LSIC	Japan-Danish EPA Inventc	2000	
23	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	without S9				Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
24	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	with S9			rat	LSIC	Japan-Danish EPA Inventc	2000	
25	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian chi	without S9				Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
26	Negative	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian cell micronucleus test					Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
27	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian cell micronucleus test					Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
28	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian cell micronucleus test					Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
29	Positive	Chromosome aberr	Chromosome aberration	in vitro	in vitro mammalian cell micronucleus test					Kirkland et a	Mutation Research 587	2005	CPDB, NTP, IARC
30													

Import of database: Horizontal import procedure

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.*xlsx*)(5);
6. Click on *Open* (6).



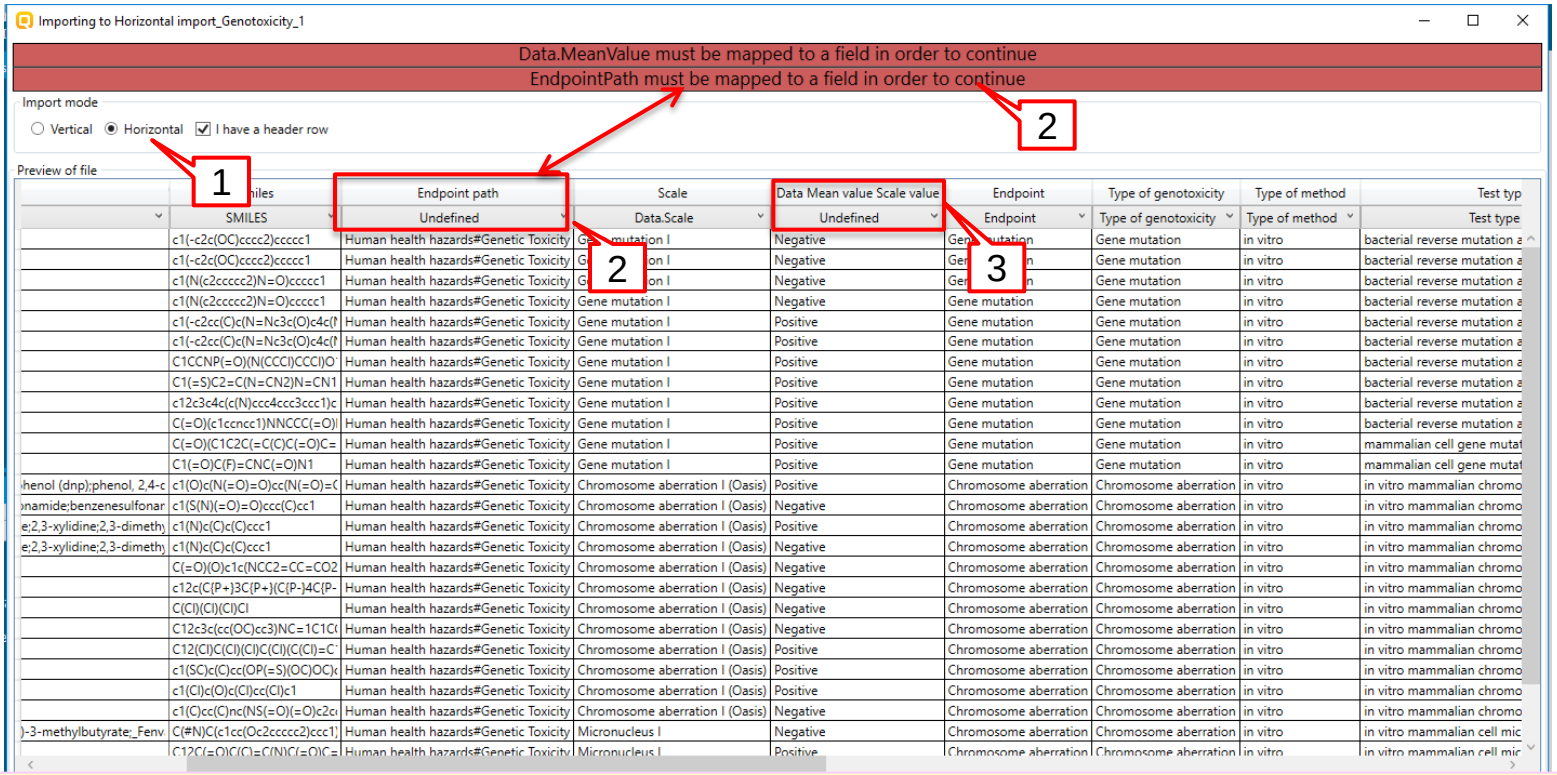
Import of database: Horizontal import procedure



CAS	Chemical name	Smiles	Endpoint path	Scale	Data Mean value	Endpoint	Type of genotoxicity	Type of method	
86260	1,1'-biphenyl,_2-methoxy-	c1(-c2c(OC)cccc2)ccccc1	Human health hazards#Genetic Toxicity	Gene mutation I	Negative	Gene mutation	Gene mutation	in vitro	bacterial rev
86260	1,1'-biphenyl,_2-methoxy-	c1(-c2c(OC)cccc2)ccccc1	Human health hazards#Genetic Toxicity	Gene mutation I	Negative	Gene mutation	Gene mutation	in vitro	bacterial rev
86306	N-NITROSODIPHENYLAMI	c1(N(c2ccccc2)N=O)ccccc1	Human health hazards#Genetic Toxicity	Gene mutation I	Negative	Gene mutation	Gene mutation	in vitro	bacterial rev
86306	N-NITROSODIPHENYLAMI	c1(N(c2ccccc2)N=O)ccccc1	Human health hazards#Genetic Toxicity	Gene mutation I	Negative	Gene mutation	Gene mutation	in vitro	bacterial rev
72571	Trypan_Blue	c1(-c2cc(C)c(N=Nc3c(O)c4c	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	bacterial rev
72571	Trypan_Blue	c1(-c2cc(C)c(N=Nc3c(O)c4c	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	bacterial rev
50180	CYCLOPHOSPHAMIDE	C1CCNP(=O)(N(CCCl)CCCl)C	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	bacterial rev
50442	Purine-6-thiol	C1(=S)C2=C(N=CN2)N=CN	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	bacterial rev
1606673	1-PYRENAMINE	c12c3c4c(c(N)ccc4ccc3ccc1)	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	bacterial rev
51127	nialamide	C(=O)(c1ccncc1)NNCCC(=O	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	bacterial rev
50760	Actinomycin_D	C(=O)(C1C2C(=C(C)C(=O)C-	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	mammalian
51218	FLUOROURACIL	C1(=O)C(F)=CNC(=O)N1	Human health hazards#Genetic Toxicity	Gene mutation I	Positive	Gene mutation	Gene mutation	in vitro	mammalian
51285	2,4-dinitrophenol;2,4-dinit	c1(O)c(N(=O)=O)cc(N(=O)=	Human health hazards#Genetic Toxicity	Chromosome aberratic	Positive	Chromosome aberration	Chromosome aberration	in vitro	in vitro mam
70553	4-methylbenzenesulfonam	c1(S(N)(=O)=O)ccc(C)cc1	Human health hazards#Genetic Toxicity	Chromosome aberratic	Negative	Chromosome aberration	Chromosome aberration	in vitro	in vitro mam
87592	2,3-dimethylaniline;2,3-dir	c1(N)c(C)c(C)ccc1	Human health hazards#Genetic Toxicity	Chromosome aberratic	Positive	Chromosome aberration	Chromosome aberration	in vitro	in vitro mam
87592	2,3-dimethylaniline;2,3-dir	c1(N)c(C)c(C)ccc1	Human health hazards#Genetic Toxicity	Chromosome aberratic	Negative	Chromosome aberration	Chromosome aberration	in vitro	in vitro mam
54319	FUROSEMIDE	C(=O)(O)c1c(NCC2=CC=CO	Human health hazards#Genetic Toxicity	Chromosome aberratic	Negative	Chromosome aberration	Chromosome aberration	in vitro	in vitro mam
50282	Estradiol_[USAN:INN]	c12c(C{P+}3C{P+}(C{P-}4C{P	Human health hazards#Genetic Toxicity	Chromosome aberratic	Negative	Chromosome aberration	Chromosome aberration	in vitro	in vitro mam

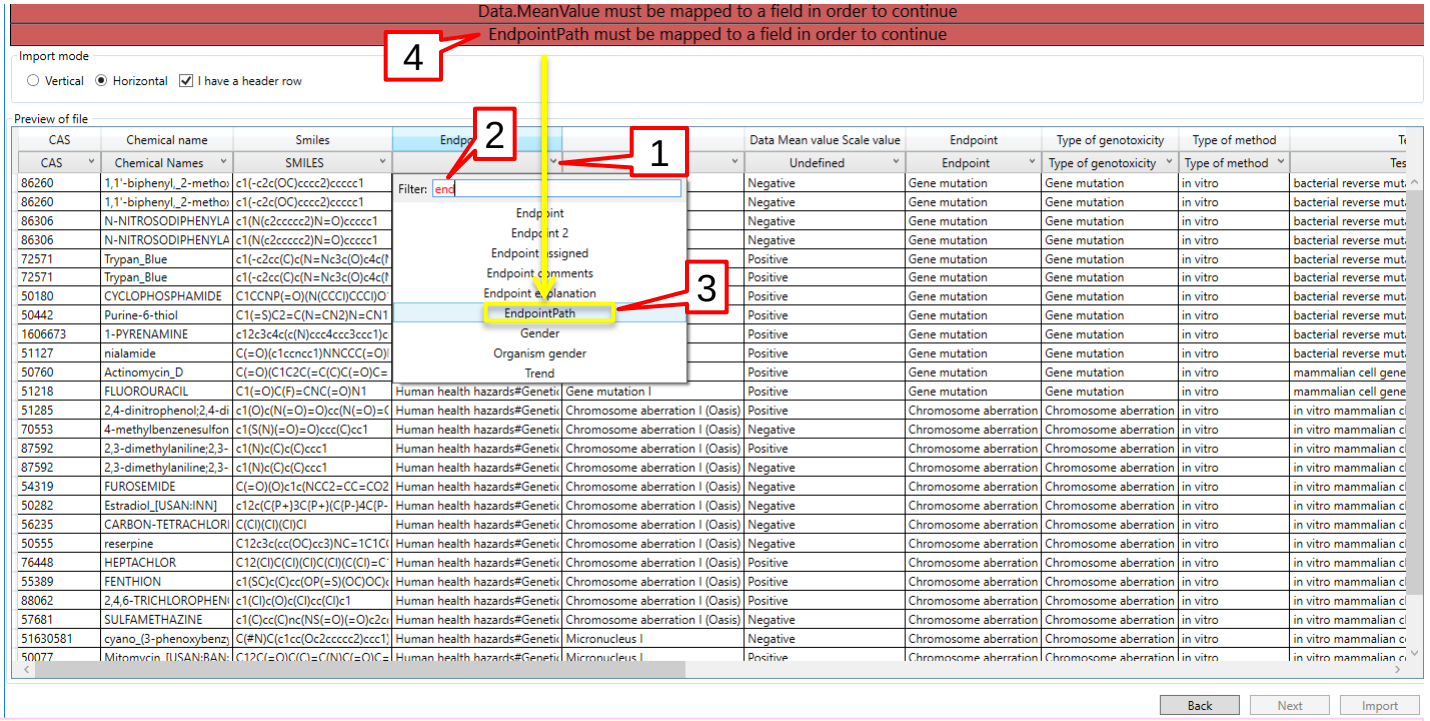
1. A preview of the imported file is shown (1);
2. Click on *Next* (2);

Import of database: Horizontal import procedure



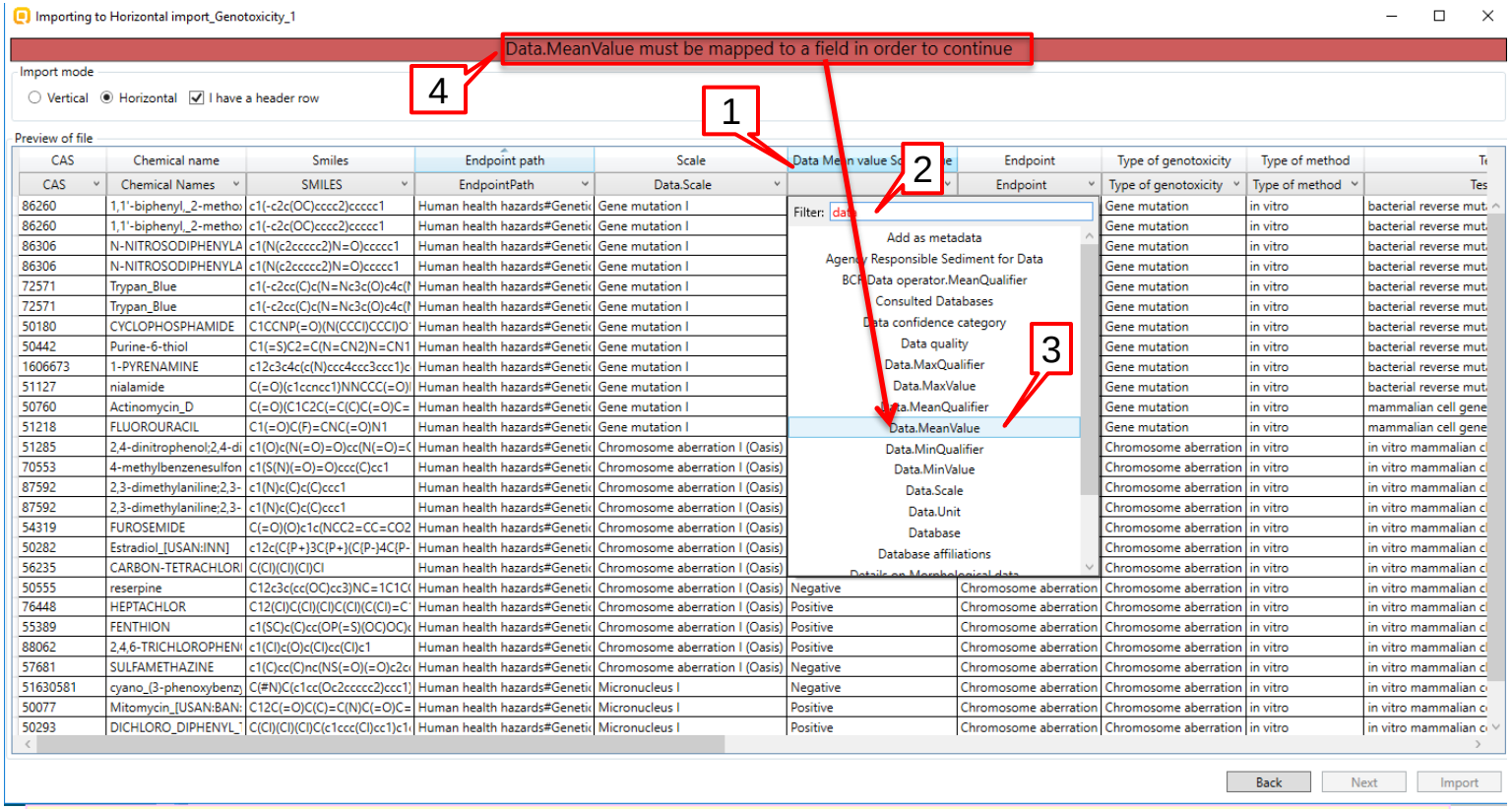
- Select *Horizontal* radio button (1)
- Column with chemical ID is automatically recognized by the system (for more details sees slide 15)
- When there are fields from the original file which cannot be mapped to the labels existing in Toolbox, burgundy colored messages are displayed on the top (2) and Undefined is written below the data which is not mapped (3). In order to map correctly the fields see next slide.

Import of database: Horizontal import procedure



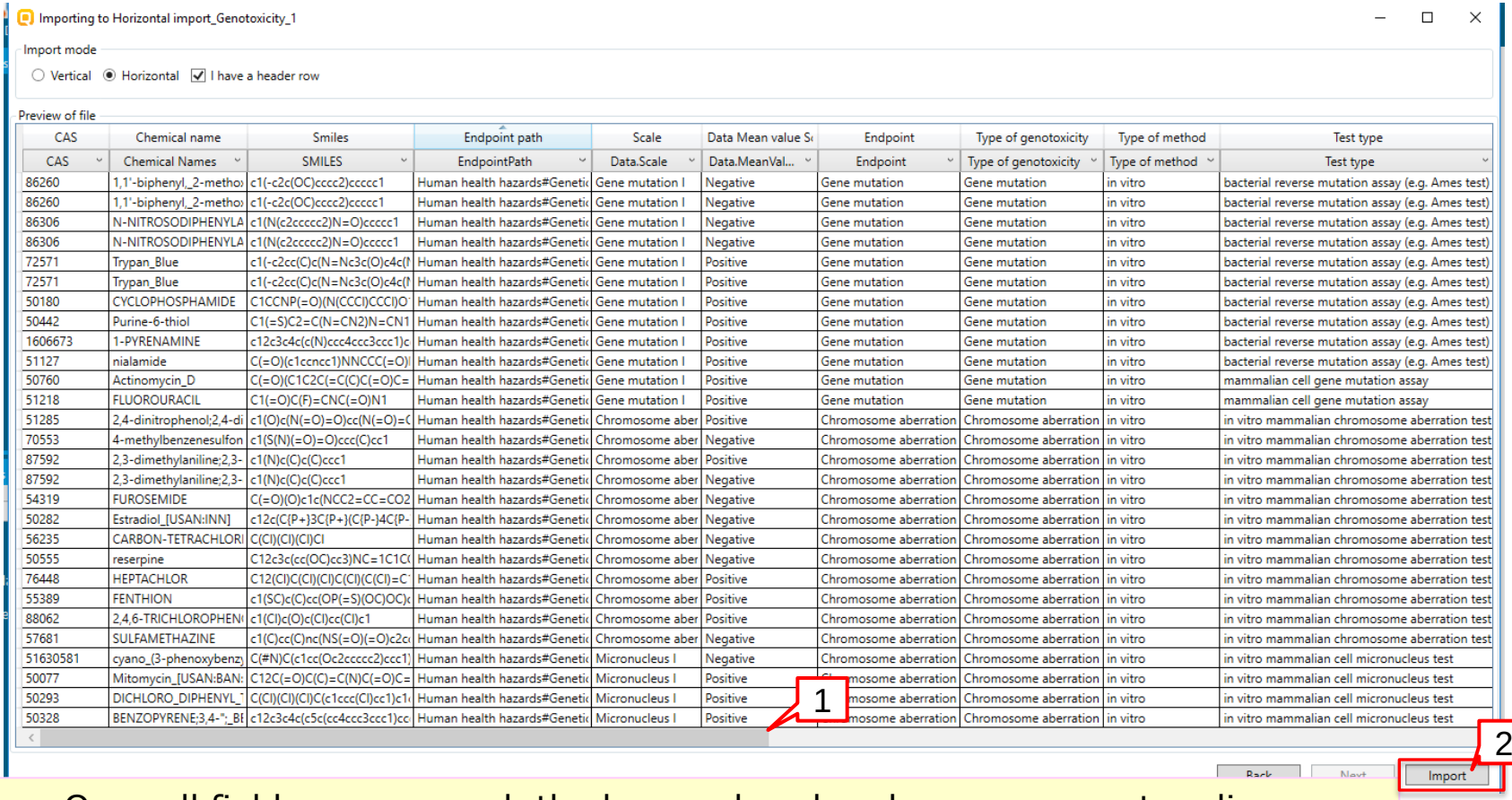
1. Open the drop-down menu (1), type in “end” in the filter (2), select EndpointPath (3).
2. The selected label has to correspond to the one in the original file and it is also written in the top message (4).

Import of database: Horizontal import procedure



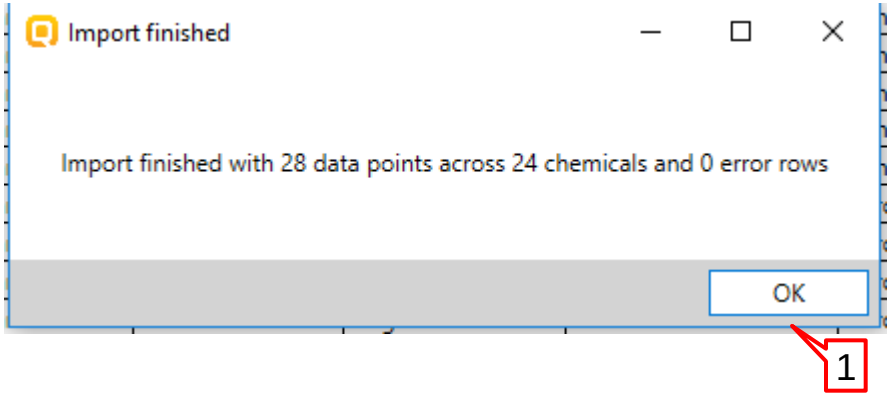
1. Open the drop-down menu (1), type in "data" in the filter (2), select Data.MeanValue (3).
2. The selected label has to correspond to the one in the original file and it is also written in the top message (4).

Import of database: Horizontal import procedure



- 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)

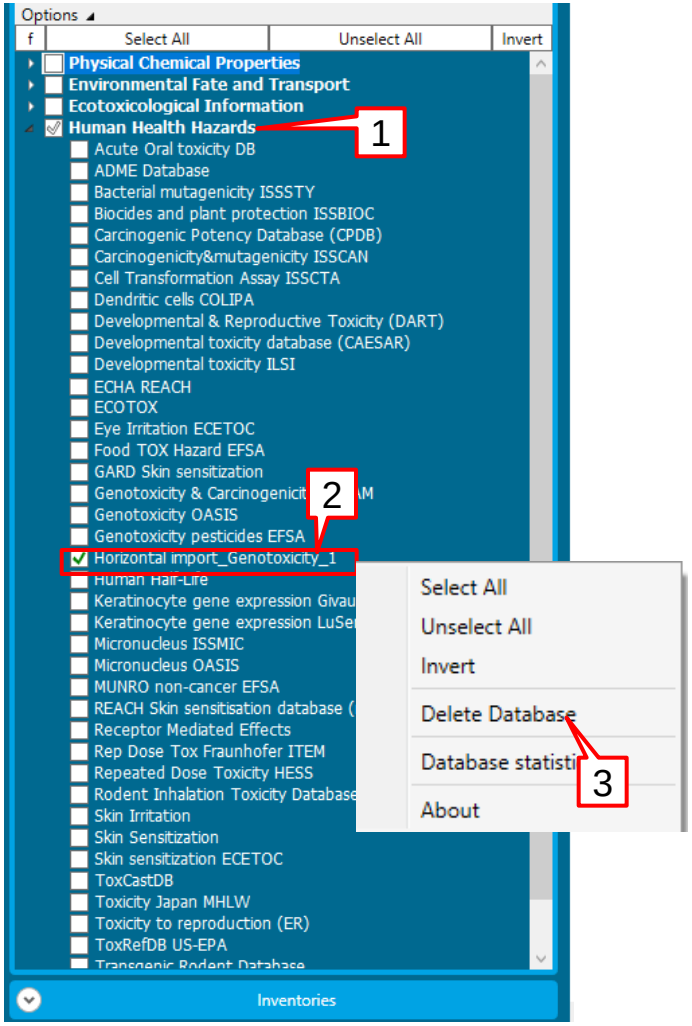
Import of database: Horizontal import procedure



- 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

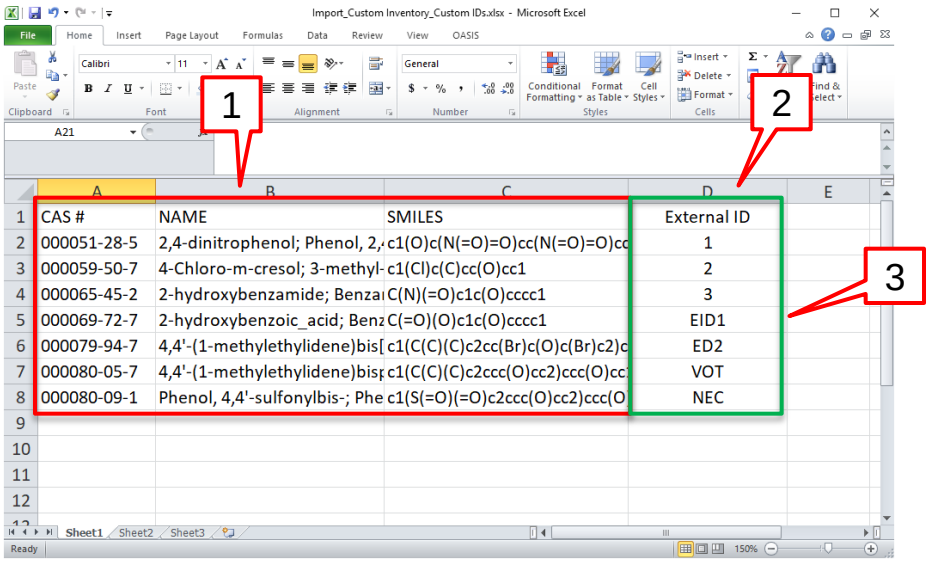
- Aim
- Keywords
- Definition of Database and Inventory
- **Import:**
  - Import of database
    - Vertical import
    - Horizontal import
  - **Import of inventory**
  - Import via IUCLID
- Export Data matrix

Import of inventory Layout

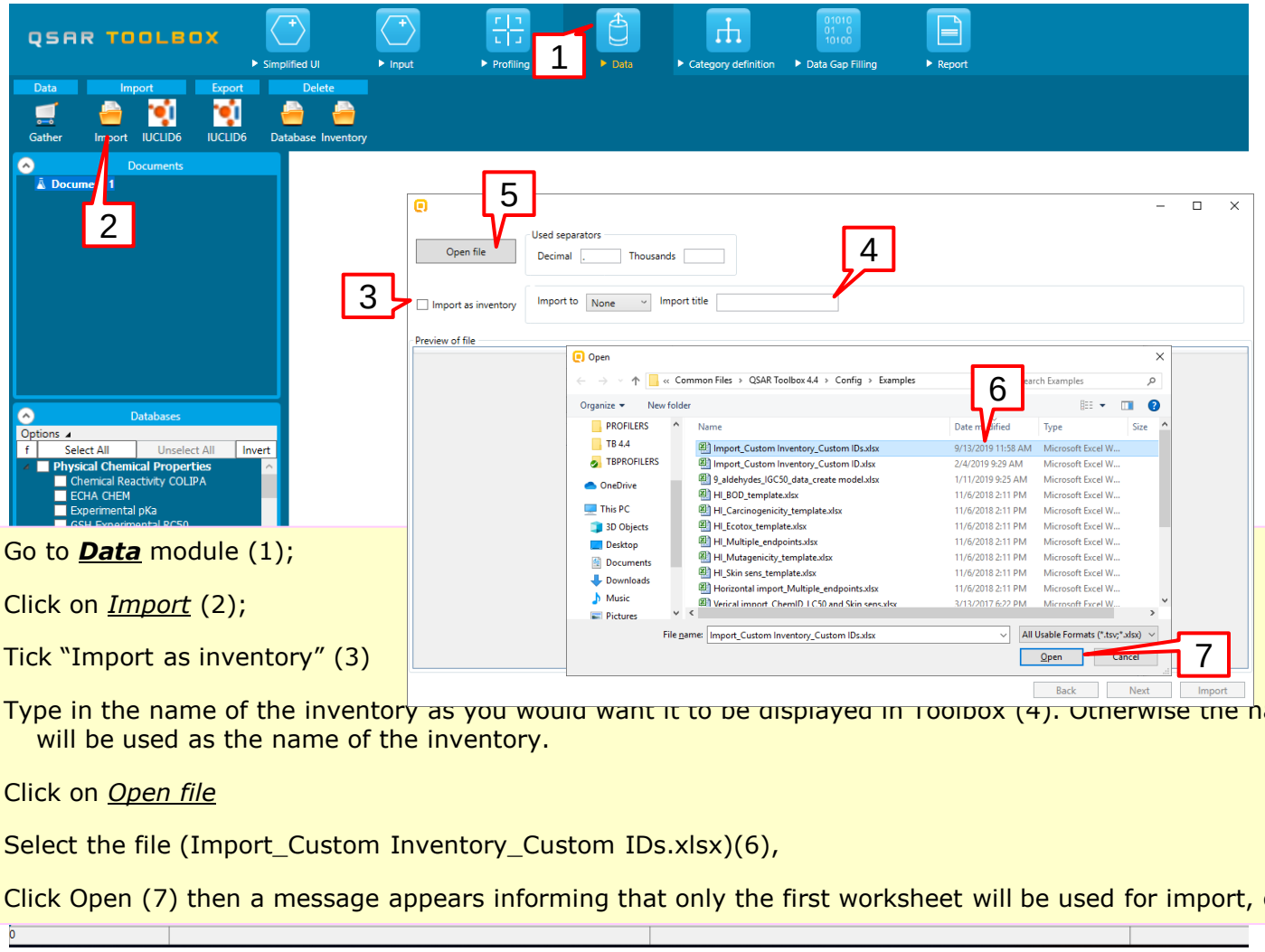
- As already presented an inventory is a collection of chemicals without any data.
- In the current example import of chemicals along with their external identifiers (custom ID) will be illustrated.

Import of inventory Layout

- 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



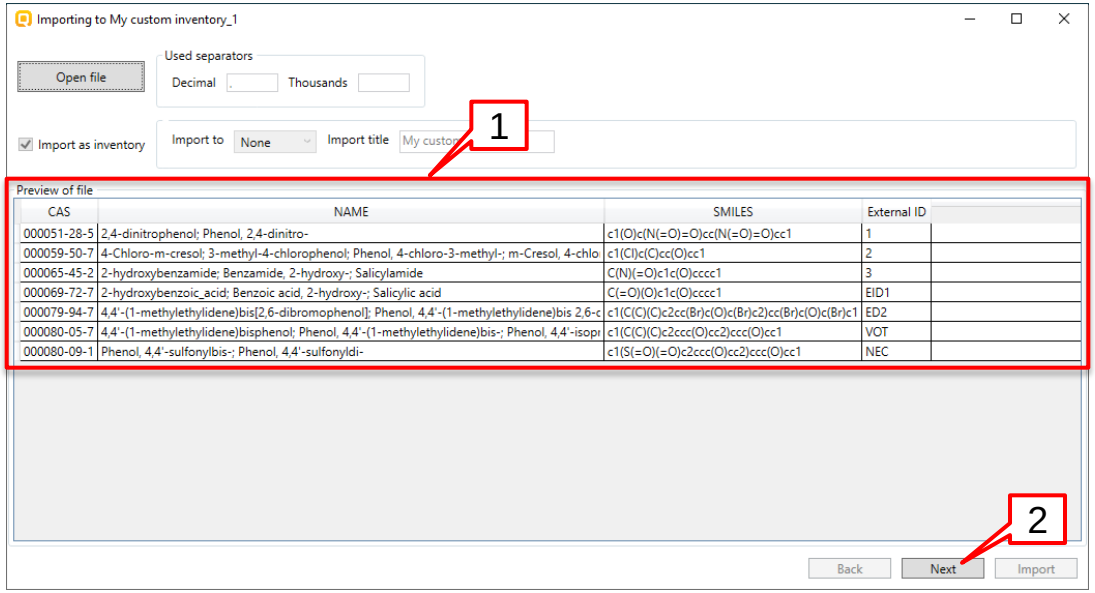
Import of inventory Procedure



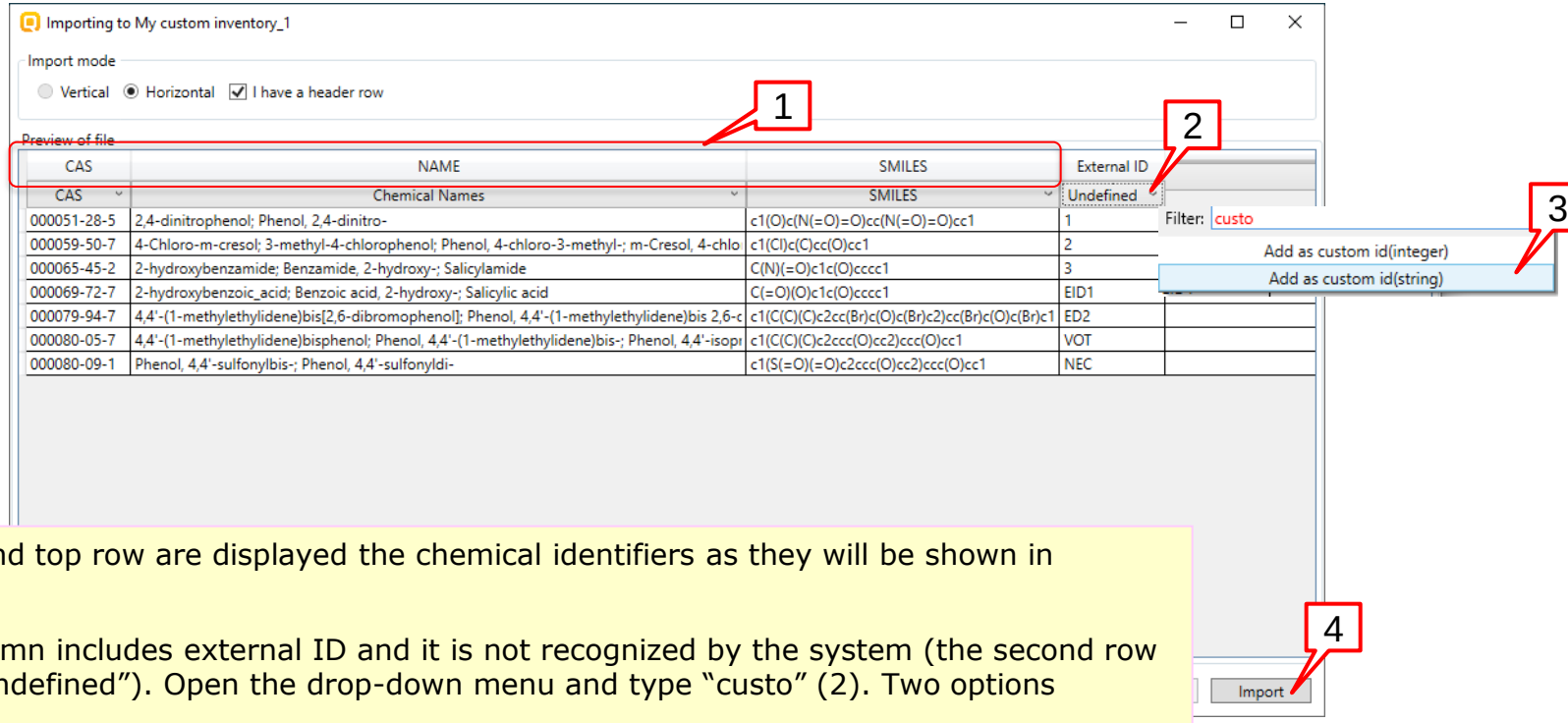
1. Go to ***Data*** module (1);
2. Click on *Import* (2);
3. Tick "Import as inventory" (3)
4. Type in the name of the inventory as you would want it to be displayed in Toolbox (4). Otherwise the name of the file will be used as the name of the inventory.
5. Click on *Open file*
6. Select the file (Import_Custom 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

Import of inventory Procedure

1. A preview of the imported file is shown (1);
2. Click on *Next* (2);

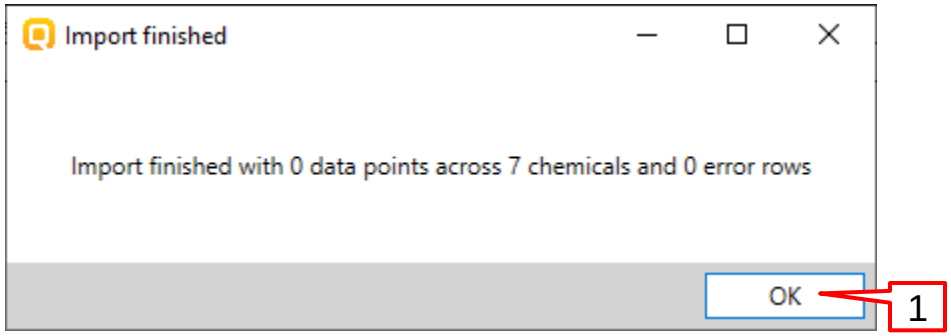


Import of inventory Procedure



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 our case external IDs are combination of letters and strings, so second option is selected here (3).
4. Click *Import* (4);

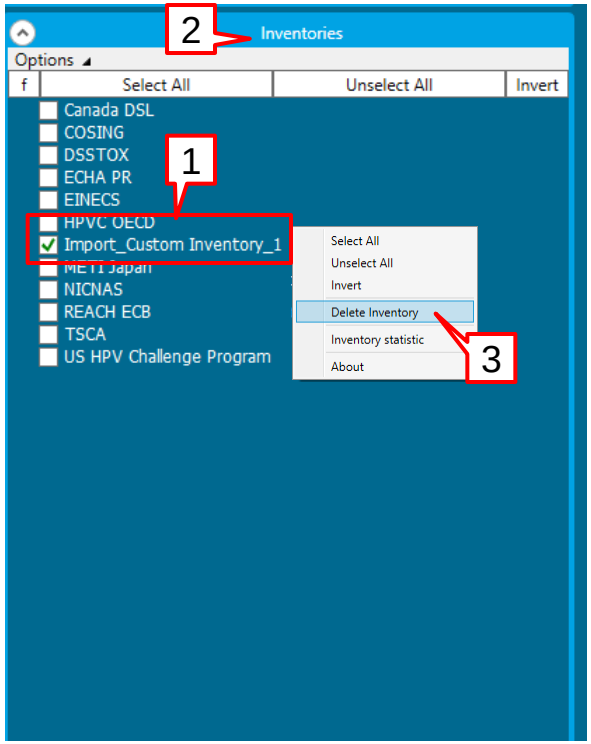
Import of inventory Procedure



- 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 inventory

- The new inventory (1) is displayed in the 'Inventories' panel (2).
- Right-click menu (3) is implemented where you can delete the inventory.
- The software automatically adds a numeration in the name ("1" in this example) in case the same inventory 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

IUCLID - background

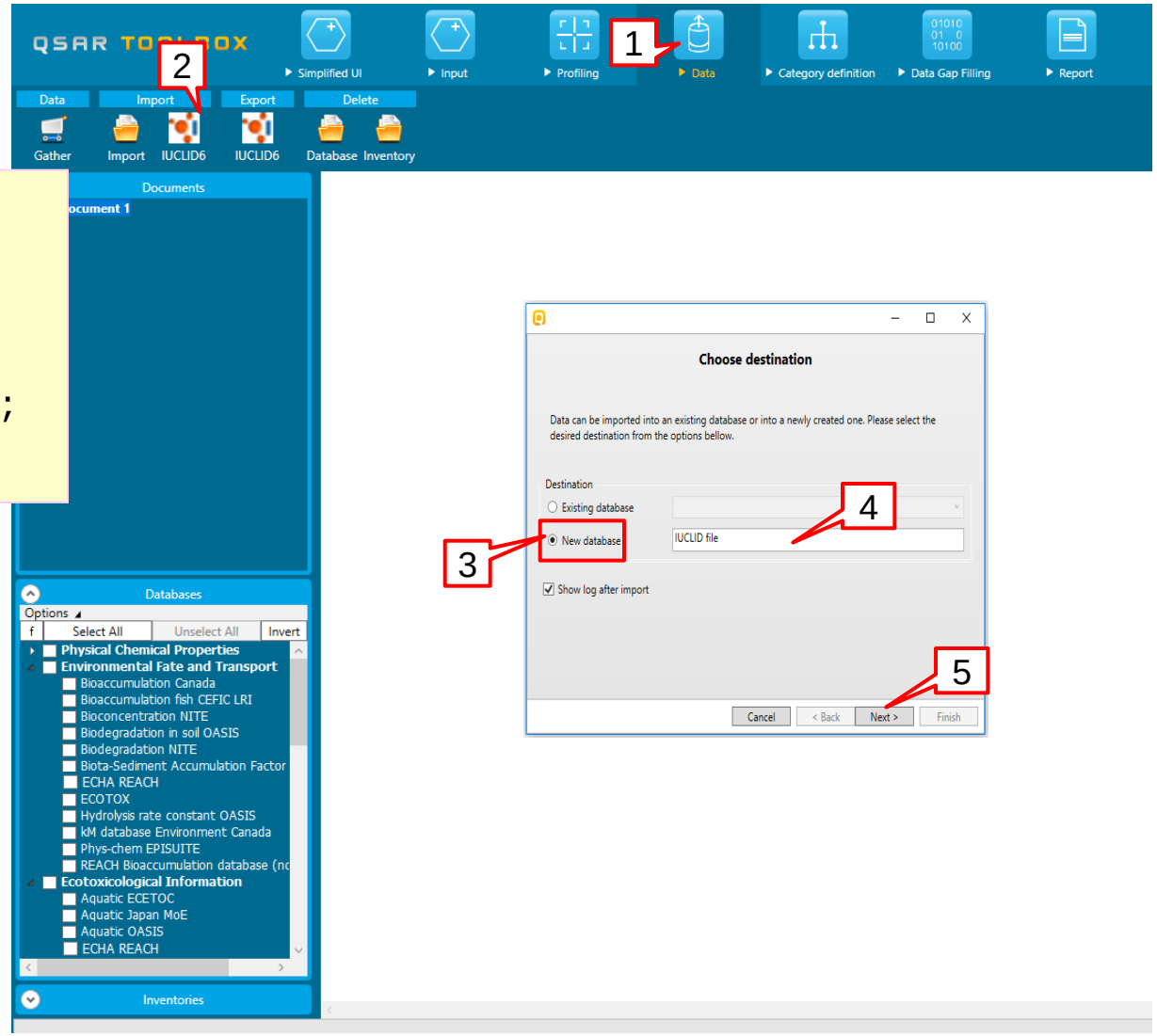
- IUCLID (International Uniform ChemicaL Information Database) is a software application to record, store, maintain and exchange data on intrinsic and hazard properties of chemical substances. It is a key software application for both regulatory bodies and the chemical industry where it is used in the implementation of various regulatory programmes

- QSAR Toolbox is connected to IUCLID (IU) through WebServices. Experimental data is allowed to be imported from IU to Toolbox only.

- Prediction obtained in the Toolbox could also be exported to IU. An example is shown in tutorial:
  Predicting aquatic toxicity to daphnia by trend analysis.pdf

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).



Note: Importing data from IUCLID requires:

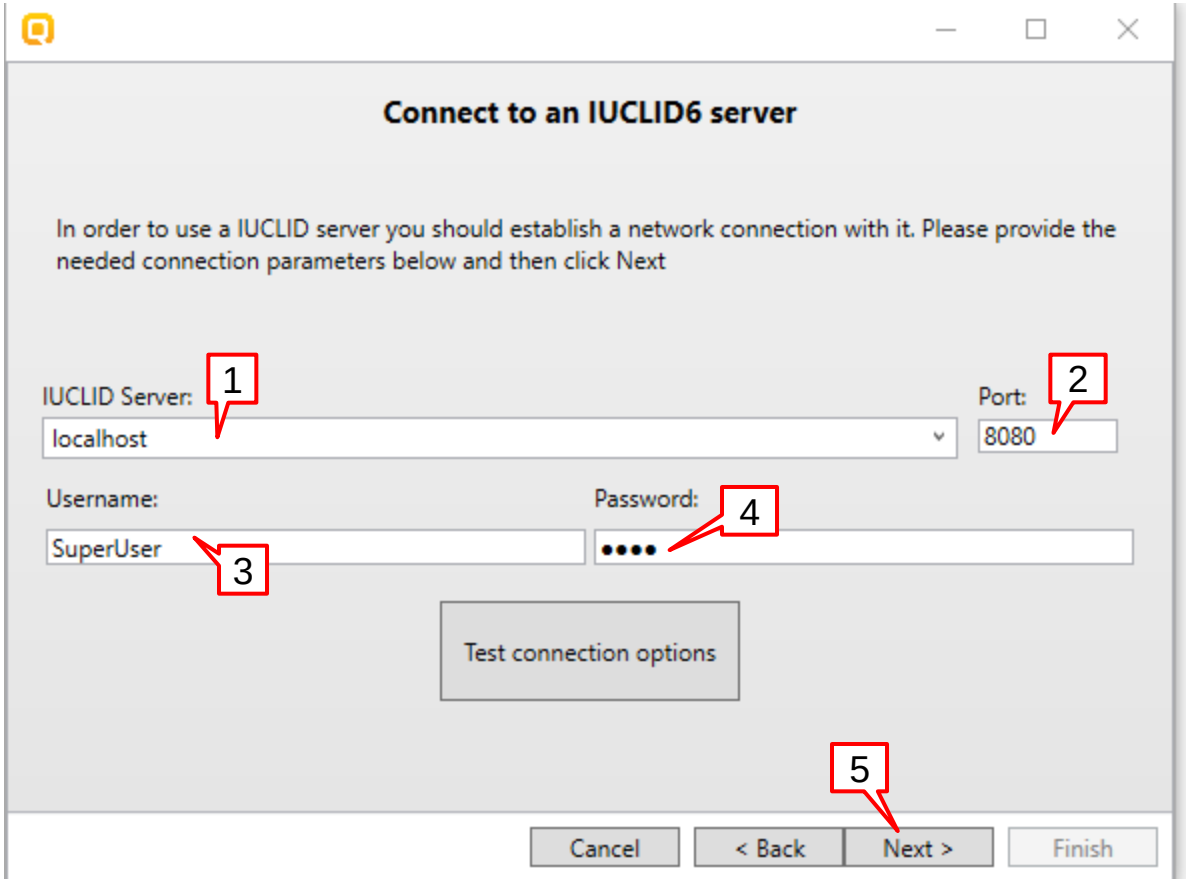
1) To have access to IU 6.3 (or 6.4)
2) IU web services to be running
3) To have a substance in the running IU server

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).

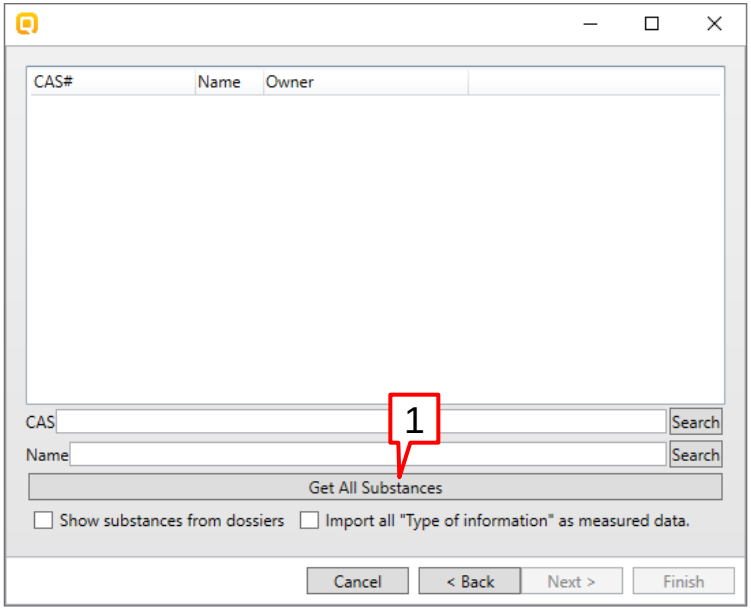
Note:

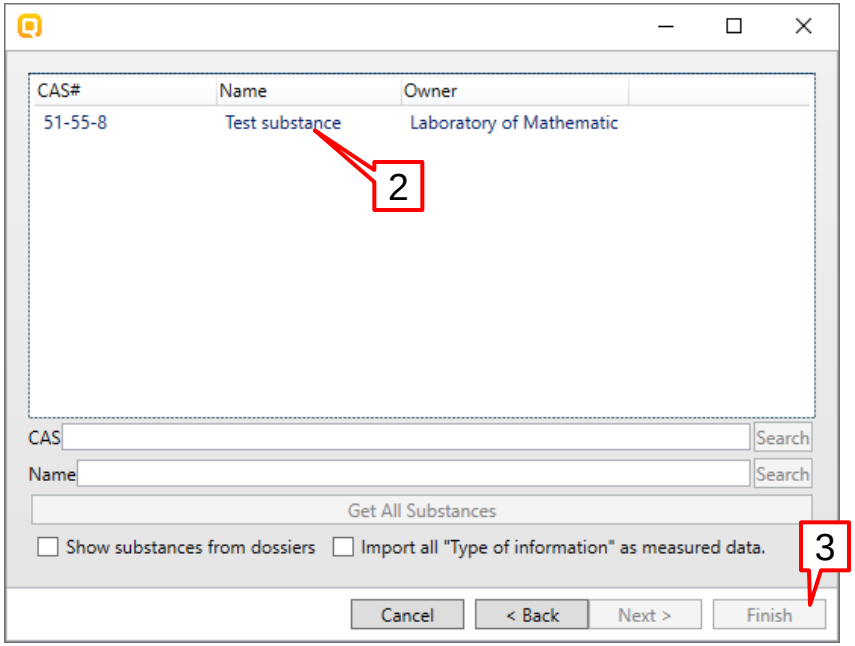
In case you don't know your IUCLID account, please ask your administrator.



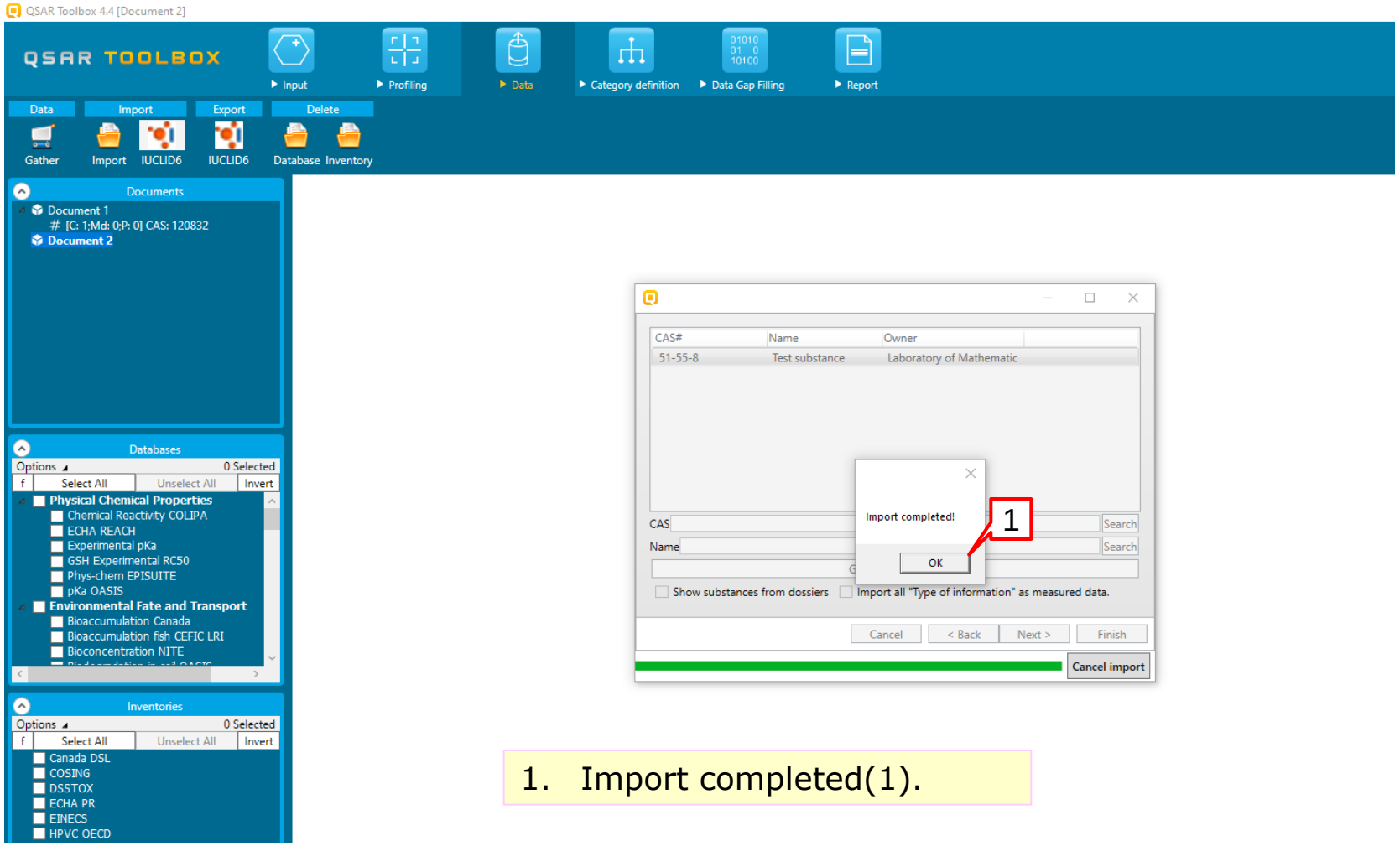
Import of data from IUCLID to Toolbox

1. Click Get All Substances (1).
2. Select the substance which you would like to import in the Toolbox (2)
3. Finish (3)



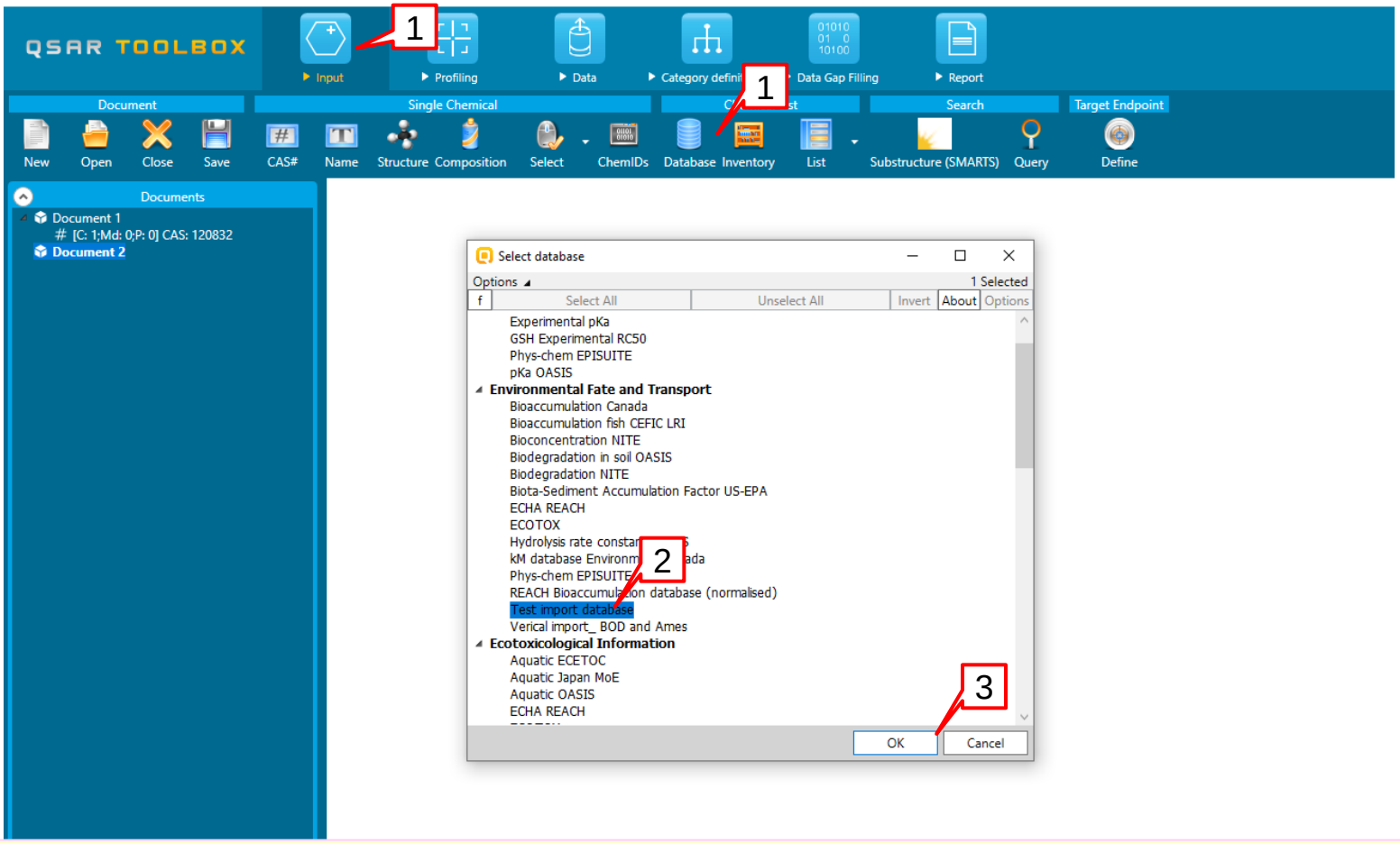


Import of data from IUCLID to Toolbox



1. Import completed(1).

Import of data from IUCLID to Toolbox



1. Click on ***Input*** module, then click on *Database* button (1)
2. Select the imported IUCLID database (in our case named Test import database) 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

Data matrix export:

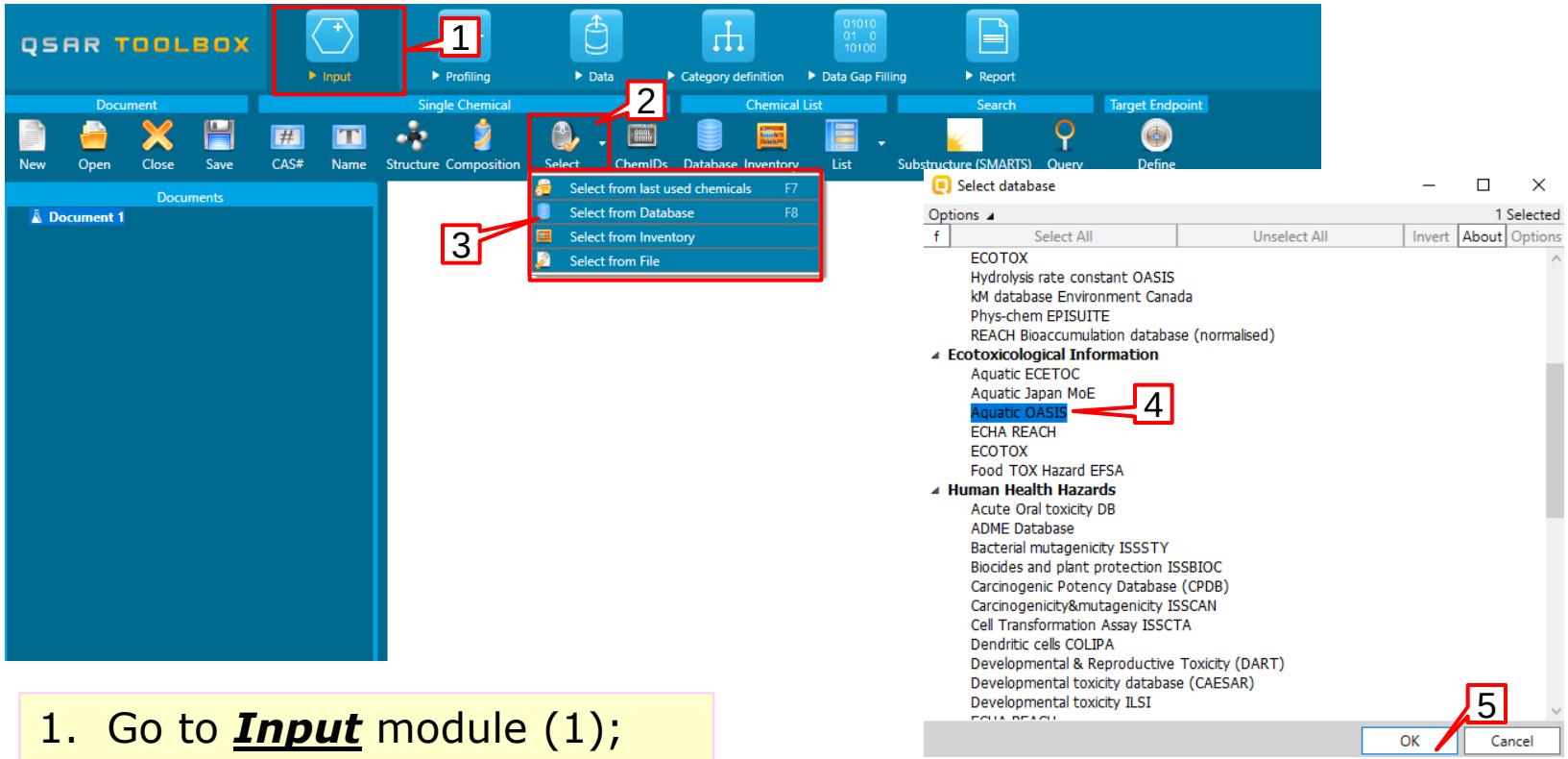
Allows to export chemicals with data and supporting information (e.g. profiling results, 2D/3D parameters, molecular formula, etc.) available on the Data matrix in text format organized in a horizontal layout.

Two options to export:

- Export data from a row
- Export whole data matrix

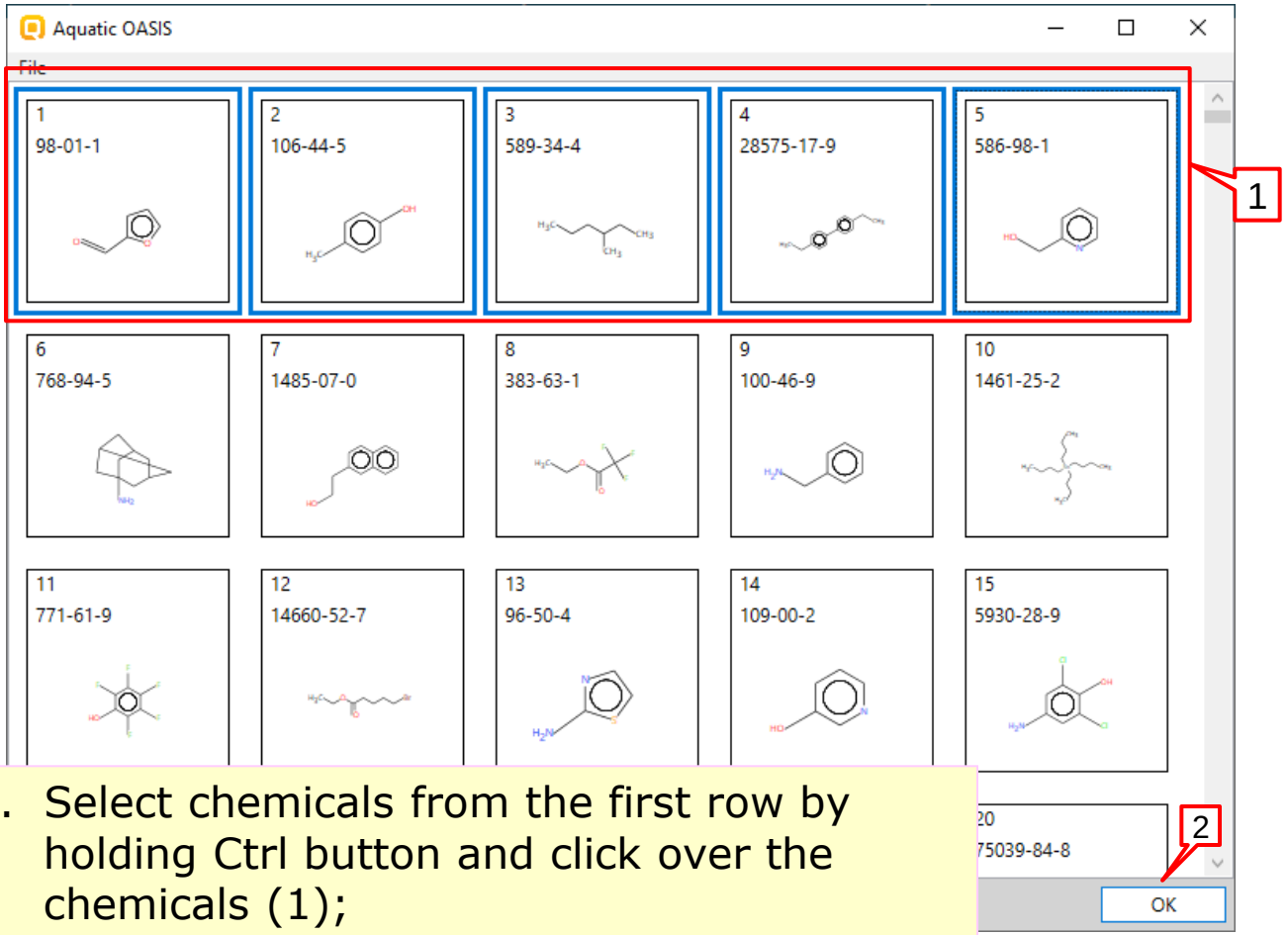
In this tutorial only the export of the **whole data matrix** is shown as it encompasses the first option as well.

Data matrix Export: Collecting data



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).

Data matrix Export: Collecting data

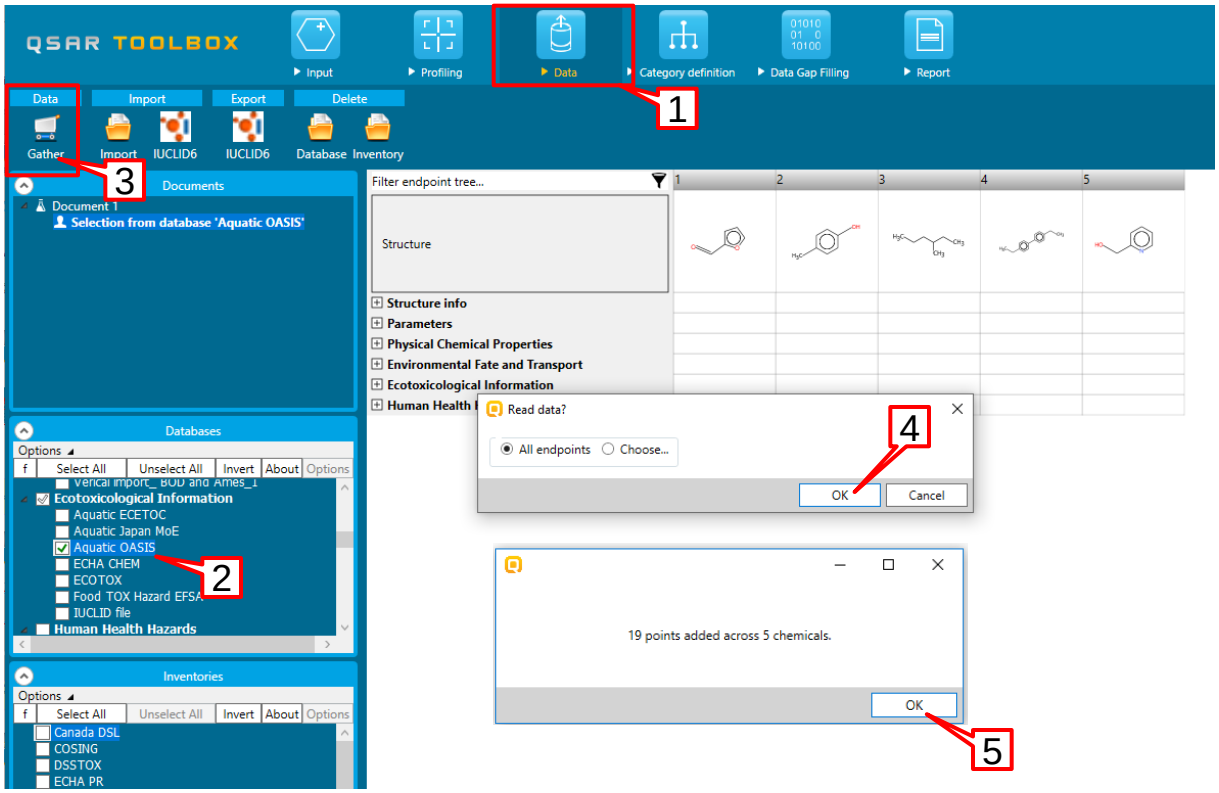


1. Select chemicals from the first row by holding Ctrl button and click over the chemicals (1);
2. Click on *OK*(2).

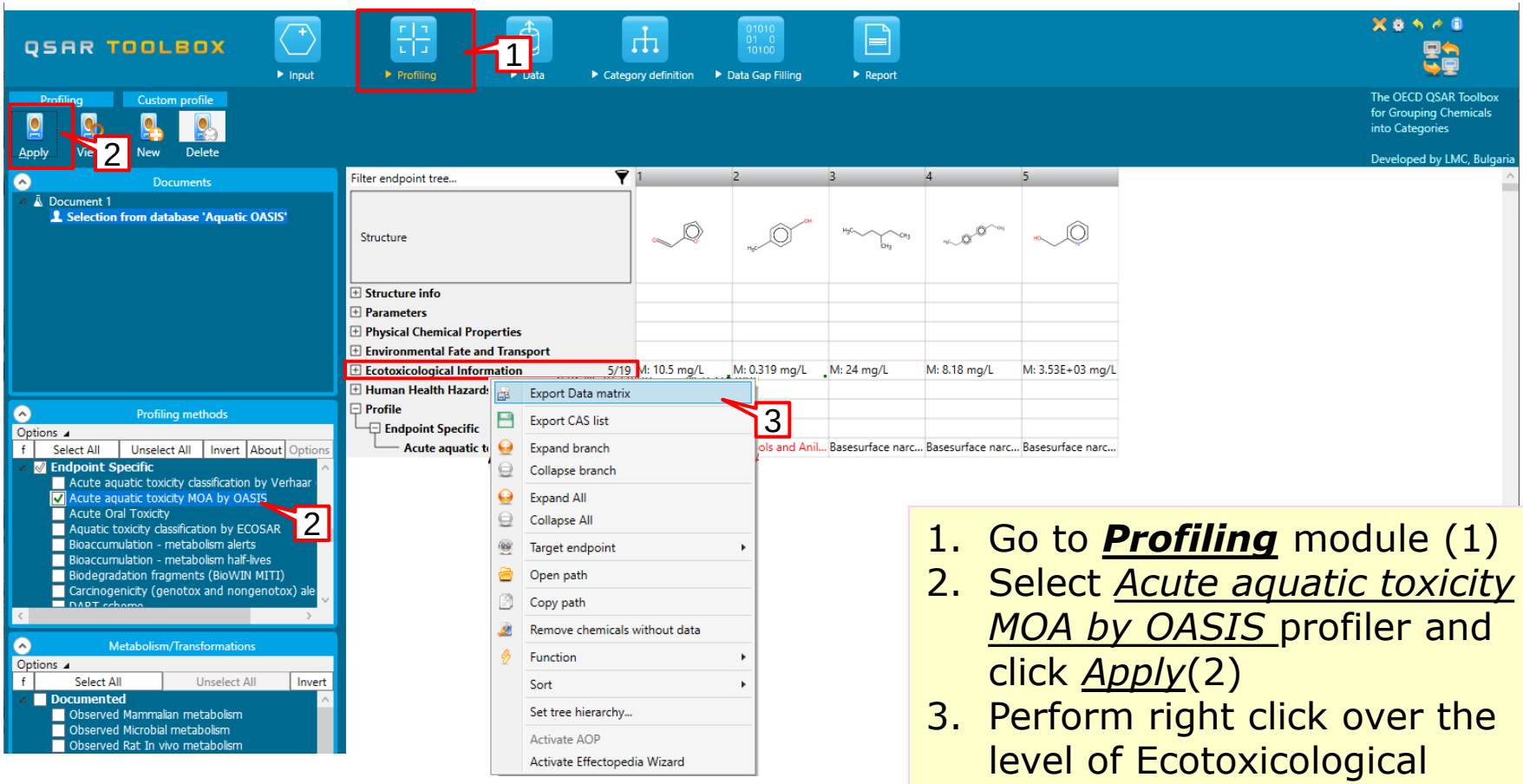
Data matrix Export: Collecting data

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).



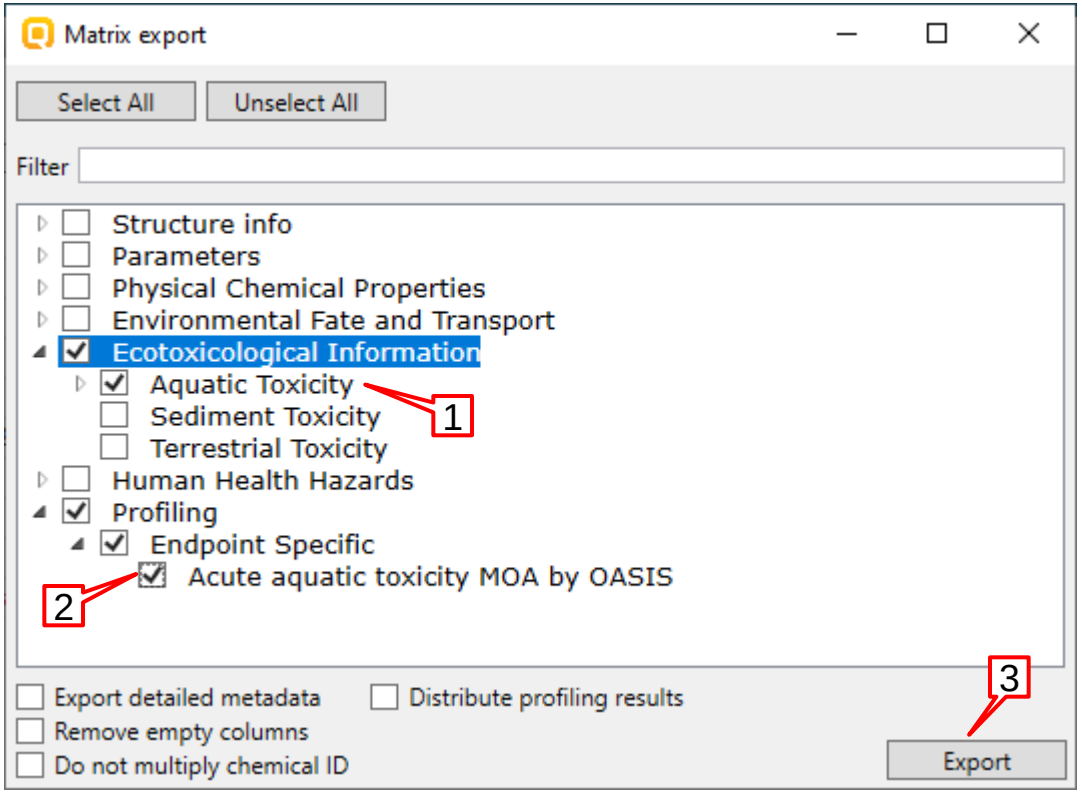
Data matrix Export



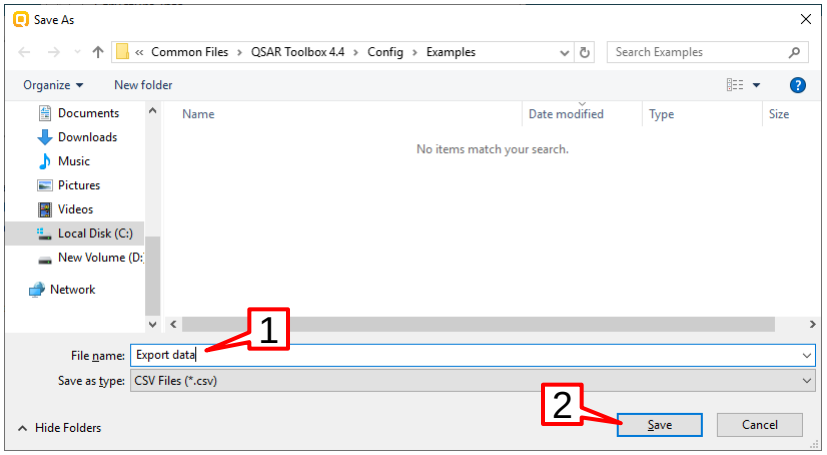
1. Go to ***Profiling*** module (1)
2. Select *Acute aquatic toxicity MOA by OASIS* profiler and click *Apply*(2)
3. Perform right click over the level of Ecotoxicological Information and select *Export data matrix* (3)

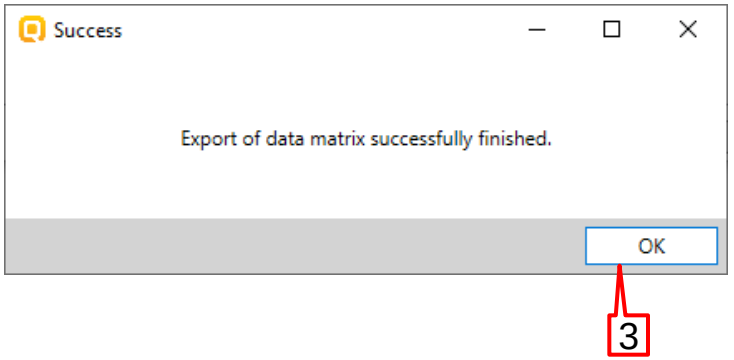
Data matrix Export

1. Keep the Aquatic toxicity level selected (1); Unselect Sediment and Terrestrial toxicity (not shown)
2. Expand the Profiling level (2) and then select Acute aquatic toxicity MOA by OASIS;
3. Click on *Export* (3).



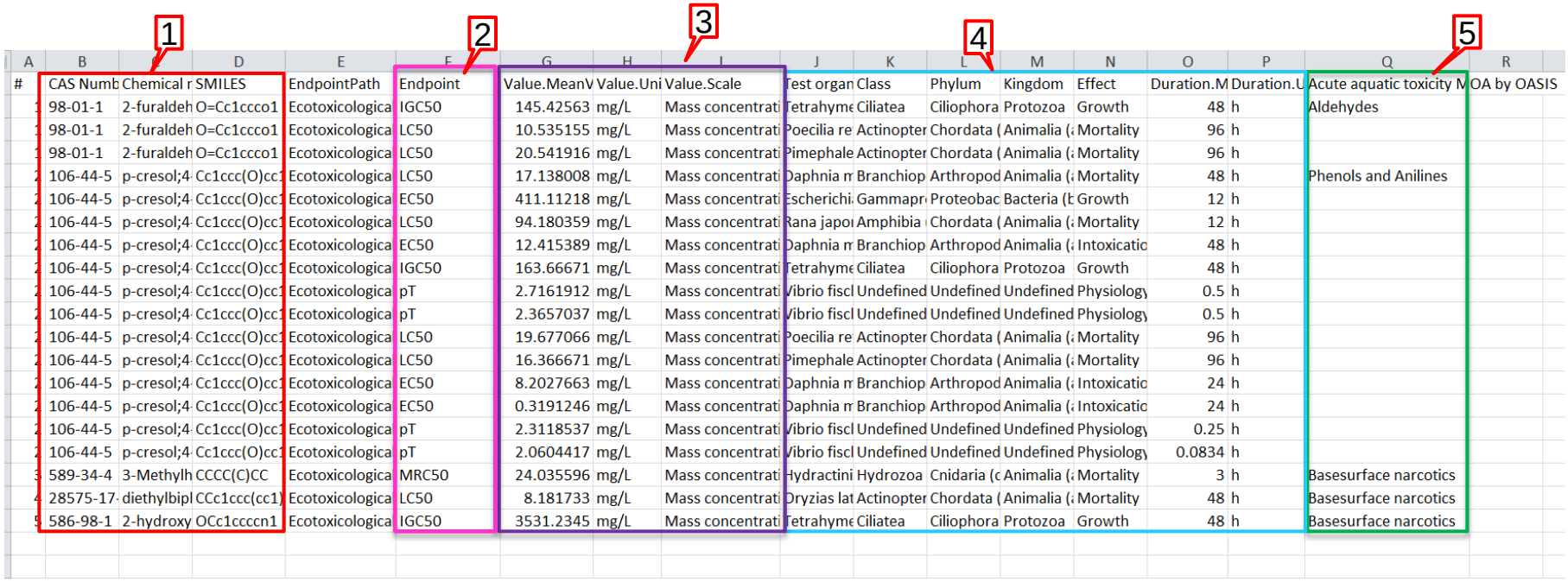
Data matrix Export





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).

Horizontal export



#	CAS Numb	Chemical r	SMILES	EndpointPath	Endpoint	Value.MeanV	Value.Uni	Value.Scale	Test organ	Class	Phylum	Kingdom	Effect	Duration.M	Duration.U	Acute aquatic toxicity M	OA by OASIS
1	98-01-1	2-furaldeh	O=Cc1ccco1	Ecotoxicologica	IGC50	145.42563	mg/L	Mass concentrati	Tetrahyme	Ciliatea	Ciliophora	Protozoa	Growth	48	h	Aldehydes	
1	98-01-1	2-furaldeh	O=Cc1ccco1	Ecotoxicologica	LC50	10.535155	mg/L	Mass concentrati	Poecilia re	Actinopter	Chordata (	Animalia (a	Mortality	96	h		
1	98-01-1	2-furaldeh	O=Cc1ccco1	Ecotoxicologica	LC50	20.541916	mg/L	Mass concentrati	Pimephale	Actinopter	Chordata (	Animalia (a	Mortality	96	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	LC50	17.138008	mg/L	Mass concentrati	Daphnia m	Branchiop	Arthropod	Animalia (a	Mortality	48	h	Phenols and Anilines	
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	EC50	411.11218	mg/L	Mass concentrati	Escherichi	Gammapr	Proteobac	Bacteria (b	Growth	12	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	LC50	94.180359	mg/L	Mass concentrati	Rana japo	Amphibia	Chordata (	Animalia (a	Mortality	12	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	EC50	12.415389	mg/L	Mass concentrati	Daphnia m	Branchiop	Arthropod	Animalia (a	Intoxicatio	48	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	IGC50	163.66671	mg/L	Mass concentrati	Tetrahyme	Ciliatea	Ciliophora	Protozoa	Growth	48	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	pT	2.7161912	mg/L	Mass concentrati	Vibrio fisc	Undefined	Undefined	Undefined	Physiology	0.5	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	pT	2.3657037	mg/L	Mass concentrati	Vibrio fisc	Undefined	Undefined	Undefined	Physiology	0.5	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	LC50	19.677066	mg/L	Mass concentrati	Poecilia re	Actinopter	Chordata (	Animalia (a	Mortality	96	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	LC50	16.366671	mg/L	Mass concentrati	Pimephale	Actinopter	Chordata (	Animalia (a	Mortality	96	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	EC50	8.2027663	mg/L	Mass concentrati	Daphnia m	Branchiop	Arthropod	Animalia (a	Intoxicatio	24	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	EC50	0.3191246	mg/L	Mass concentrati	Daphnia m	Branchiop	Arthropod	Animalia (a	Intoxicatio	24	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	pT	2.3118537	mg/L	Mass concentrati	Vibrio fisc	Undefined	Undefined	Undefined	Physiology	0.25	h		
2	106-44-5	p-cresol;4	Cc1ccc(O)cc1	Ecotoxicologica	pT	2.0604417	mg/L	Mass concentrati	Vibrio fisc	Undefined	Undefined	Undefined	Physiology	0.0834	h		
3	589-34-4	3-Methylh	CCCC(C)CC	Ecotoxicologica	MRC50	24.035596	mg/L	Mass concentrati	Hydractini	Hydrozoa	Cnidaria (c	Animalia (a	Mortality	3	h	Basesurface narcotics	
4	28575-17-	diethylbipl	CCc1ccc(cc1)	Ecotoxicologica	LC50	8.181733	mg/L	Mass concentrati	Dryzias lat	Actinopter	Chordata (	Animalia (a	Mortality	48	h	Basesurface narcotics	
5	586-98-1	2-hydroxy	OCc1ccccn1	Ecotoxicologica	IGC50	3531.2345	mg/L	Mass concentrati	Tetrahyme	Ciliatea	Ciliophora	Protozoa	Growth	48	h	Basesurface narcotics	

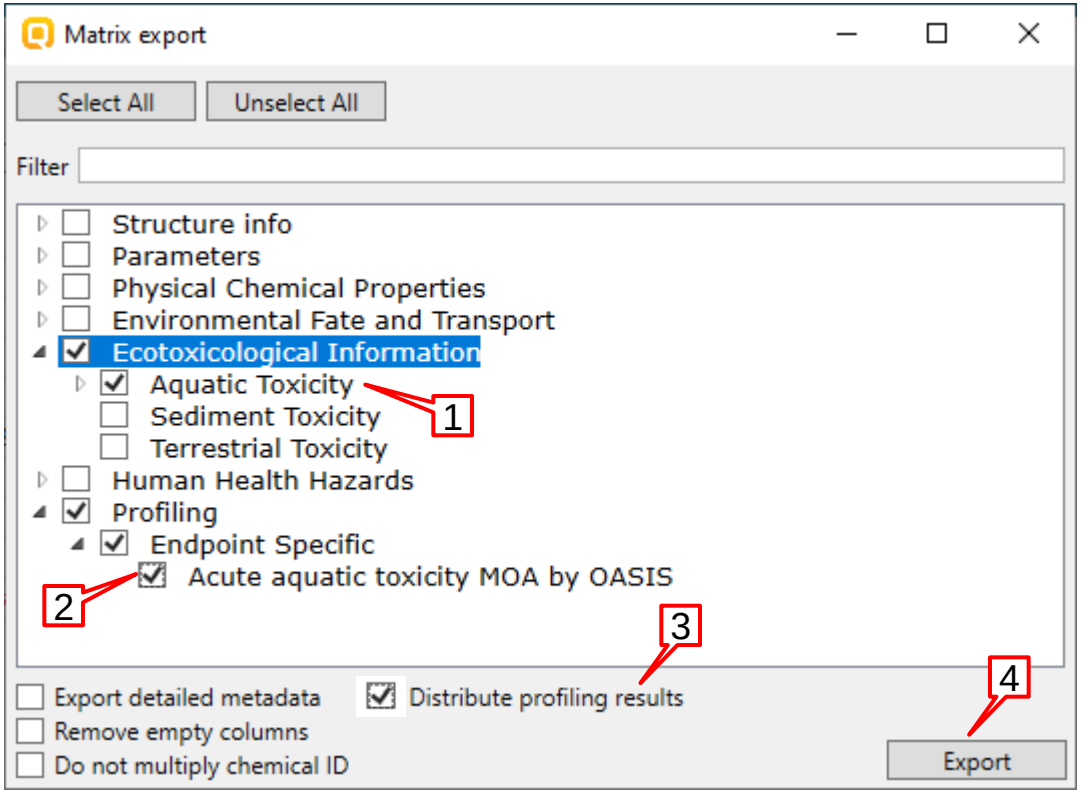
The file can be opened in excel. It contains the following main sections: *chemical identity of each chemical* (red, 1), *endpoint* (pink, 2) *experimental data and unit* (purple, 3), metadata (blue, 4), profiler results (green, 5). Here, if the chemical has several experimental data each one is listed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

Data matrix Export

Additional option to export profiling results in a format approved by EFSA

1. Keep the Aquatic toxicity level selected (1); Unselect Sediment and Terrestrial toxicity (not shown)
2. Expand the Profiling level (2) and then select Acute aquatic toxicity MOA by OASIS;
3. Check Distribute profiling results
4. Click on *Export* (3).

For more details, see next slide.



Data matrix Export

Chemicals used in the workflow

Profilers used in the screening

	A	B	C	D	E	F	G
	#	CAS Numb	SMILES	Acute aquatic toxicity MOA by OASIS	Acute aquatic toxicity MOA by OASIS	Acute aquatic toxicity MOA by OASIS	
				Aldehydes	Phenols and Anilines	Basesurface narcotics	
	1	98-01-1	O=Cc1ccco1	X			
	2	106-44-5	Cc1ccc(O)cc1		X		
	3	589-34-4	CCCC(C)CC			X	
	4	28575-17-9	CCc1ccc(cc1)-c1ccc(CC)cc1			X	
	5	586-98-1	OCc1ccccn1			X	
	0 chemical(s) were marked private and not exported.						

Label indicating that chemicals has profiling result according to the selected profiler

Categories of the profiler found for the list of chemicals

This is a new template where each chemical is listed in a separate row, while the different profiling results identified for the list of analyzed chemicals appear as captions in separate columns. Indication for the chemicals that have specific profiling results according to the applied profiler is labeled with "X" sign.

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!