QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.4.1

Step by step example on how to predict acute aquatic toxicity to Daphnia for 3-ethyl-5-methyl-3-methoxyphenol by the trend analysis approach

Outlook

- **Background**
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise

Background

- This is a step-by-step presentation designed to take the user of the Toolbox through the workflow of a data gap filling exercise by trend analysis approach.

Outlook

- Background
- **Keywords**
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise

Keywords

TARGET CHEMICAL – a 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)

PROFILER - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. organic functional groups), mechanistic (e.g. Protein binding by OECD) and endpoint-specific (e.g. in vitro in vitro mutagenicity (Ames test) alerts by ISS) profilers.

ALERT - the profilers consist of sets of rules or alerts. Each of the rules consists of a set of queries. The queries could be related to the chemical structure, physicochemical properties, experimental data, comparison with the target or list with substances and external queries from other predefined profilers (reference queries).

CATEGORY – "group" of substances sharing same characteristics (e.g. the same functional groups or mode of action). In a typical Toolbox workflow, it consists of the target chemical and its analogues gathered according to the selected profilers

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

PREDICTION – An outcome obtained for the target chemical by different gap filling approach methods (e.g. read-across, trend analysis, QSAR model predictions)

TREND ANALYSIS - The appropriate data-gap filling method for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.

IUCLID 6 (IU6) – IUCLID is a software to record, store, maintain and exchange data on intrinsic and hazard properties of chemical substances. Toolbox has a functionality to export a prediction or import of data to/from IU6

Outlook

- Background
- Keywords
- **Objectives**
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise

Objectives

- **This presentation reviews a number of functionalities of the Toolbox:**
  - Identify analogues for a target chemical
  - Retrieve experimental results available for those analogues
  - Fill data gaps by trend-analysis

Outlook

- Background
- Keywords
- Objectives
- **Specific Aims**
- Trend analysis
- The exercise
- Workflow of the exercise

Specific Aims

- To review the workflow of the Toolbox.
- To review the six modules of the Toolbox.
- To reacquaint the user with the basic functionalities within each module.
- To explain the rationale behind each step of the exercise.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- **Trend analysis**
- The exercise
- Workflow of the exercise

Trend Analysis
Overview

- For a given (eco)toxicological endpoint, the members of a category are often related by a trend (e.g. increasing, decreasing or constant). The trend could be related to molecular mass, carbon chain length, or to some other physicochemical property.

- A demonstration of consistent trends in the behaviour of a group of chemicals is one of the desirable attributes of a chemical category and one of the indicators that a common mechanism for all chemicals is involved. When some chemicals in a category have measured values and a consistent trend is observed, missing values can be estimated by simple scaling from the measured values to unmeasured values as a means of filling data gaps.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- **The exercise**
- Workflow of the exercise

Exercise

- In this exercise we will predict the acute toxicity to daphnids for an untested compound, (3-ethyl-5-methyl-4-methoxyphenol), which is the "target" chemical.
- The target endpoint is LC50, 48h, Mortality, *D.magna.*
- This prediction will be accomplished by collecting a set of test data for chemicals considered to be in the same category as the target molecule.
- The category will be defined using the following categorization schemes:

  - *Acute aquatic toxicity classification by ECOSAR* – for primary grouping.

  - *Acute aquatic toxicity MOA by OASIS* – for mechanistic refining the category.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**

Workflow

- **Toolbox has six modules which are used in a sequential workflow:**
  - Chemical Input
  - Profiling
  - Data
  - Category Definition
  - Filling Data Gaps
  - Report

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - **Chemical Input**

Chemical Input
Overview

- This module provides several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input
Ways of Entering a Chemical

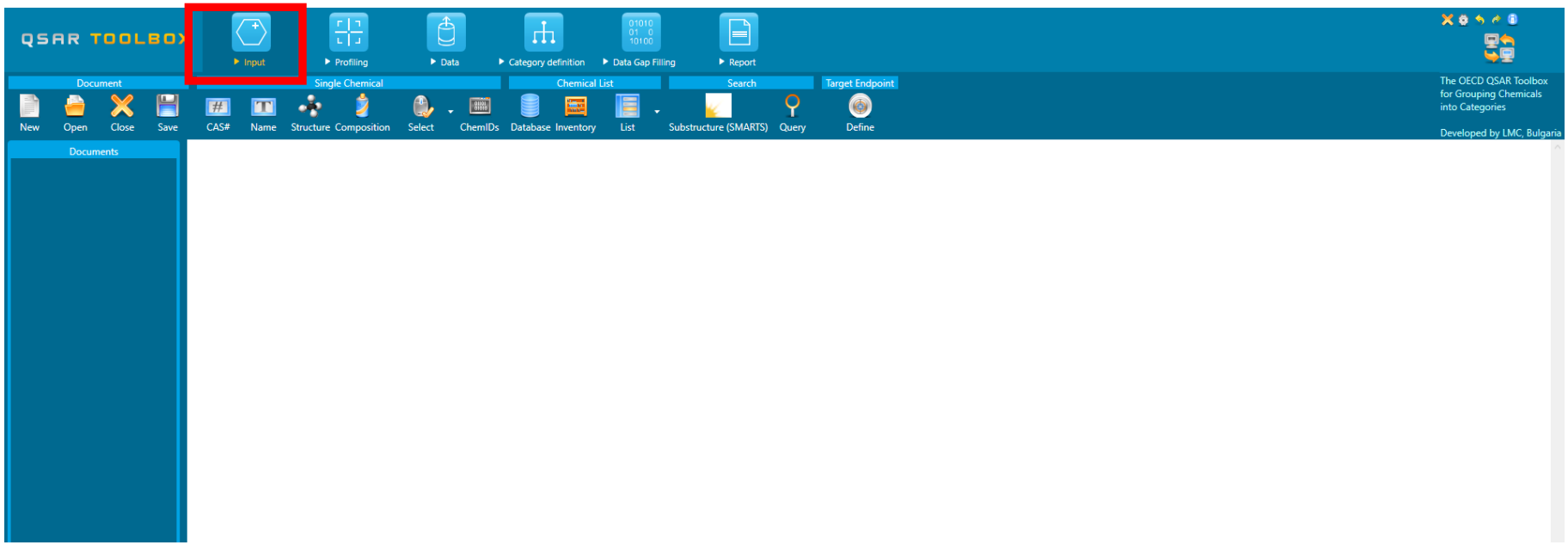
User Alternatives for Chemical ID:

A. Single target chemical
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Substructure by using SMART

B. Group of chemicals
- User List/Inventory
- Specialized Databases

Getting Started

- Open Toolbox.
- Click "Input".

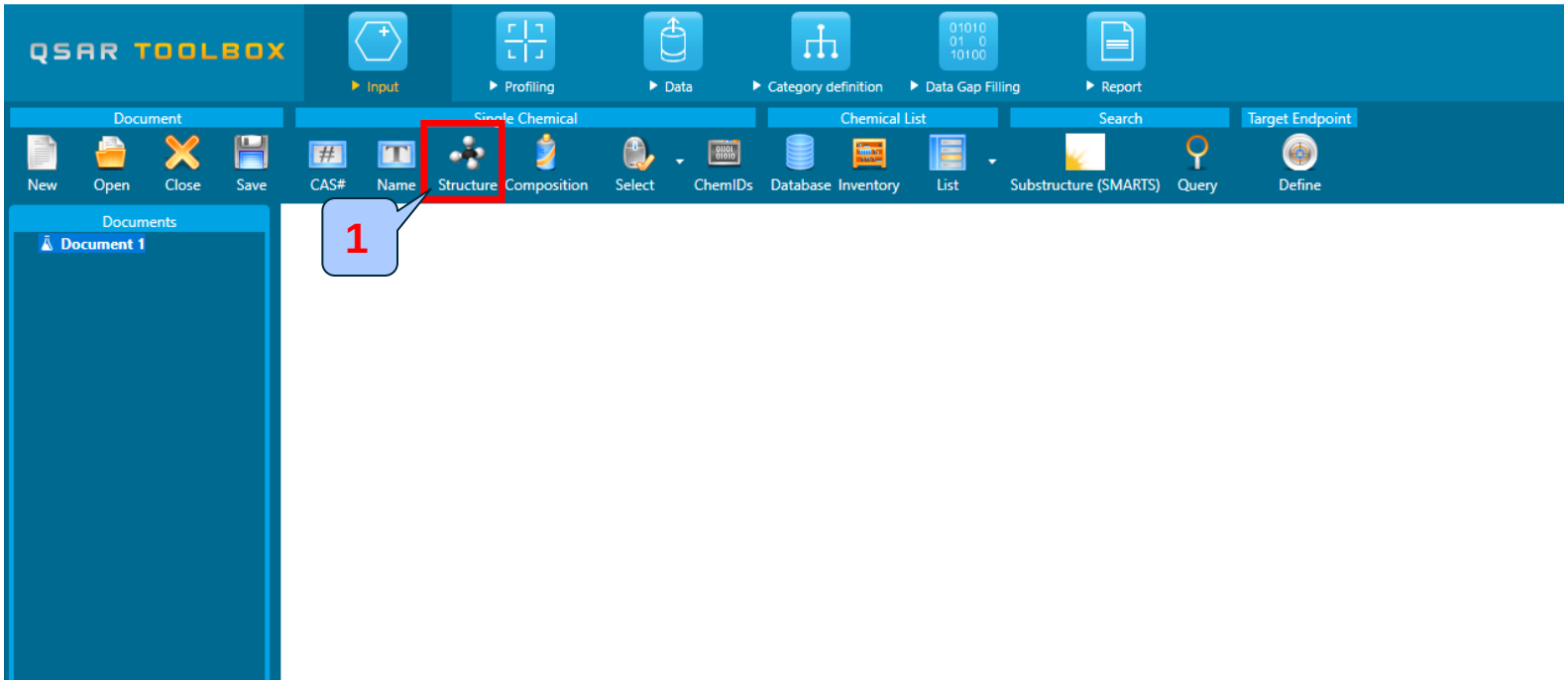


Chemical Input by Drawing

- Input of the target chemical by drawing varies in difficulty with the structural complexity of the molecule.
- It is accomplished by a series of point-click-move-click operations within the 2D-editor which drops down when you click on "structure" (see next screen shot).
- The subsequent series of slides will take you through the process for inputting the target chemical.

Chemical Input Screen

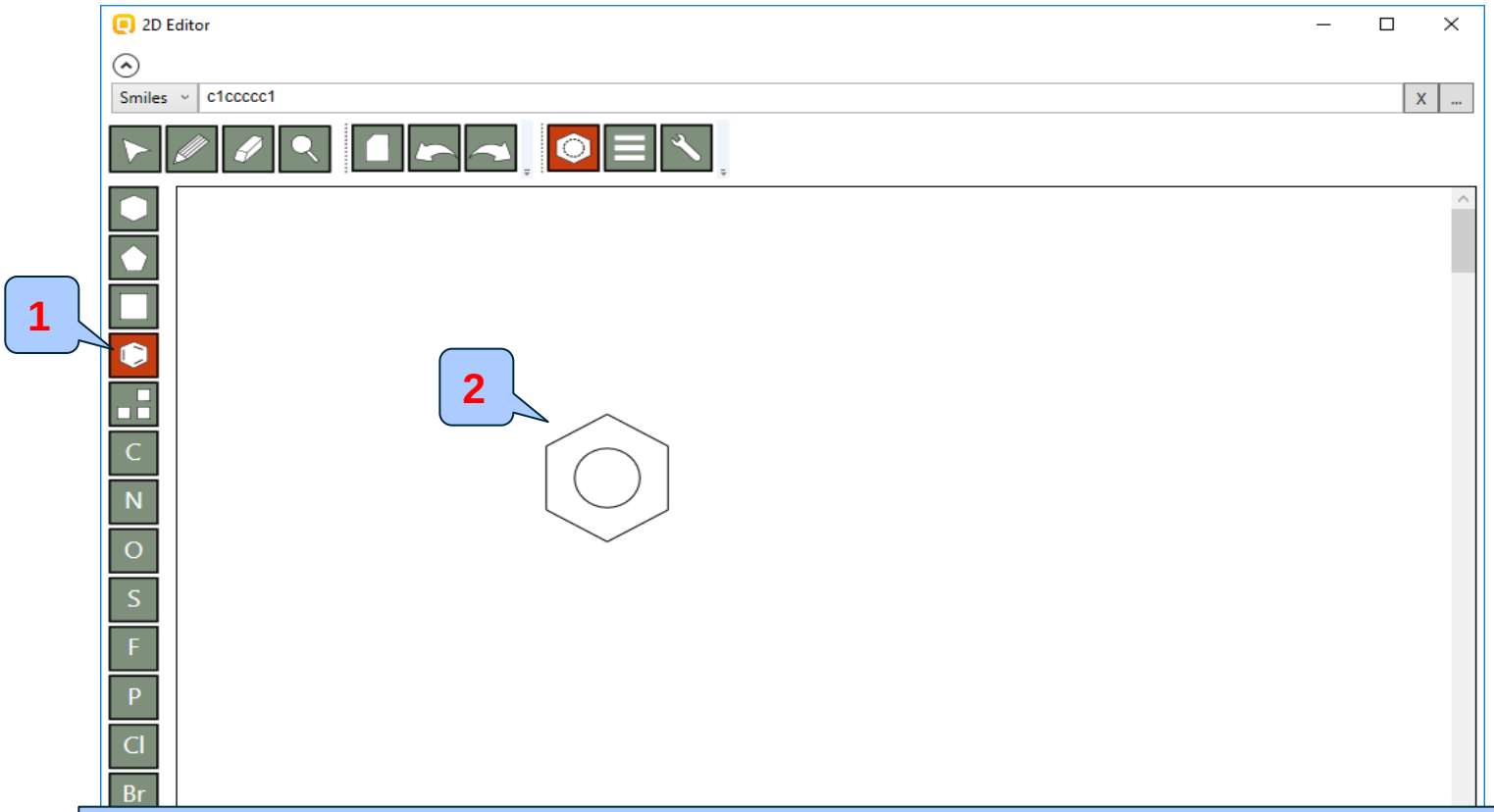
Input target chemical by drawing



1. Click **Structure** button.

Chemical Input

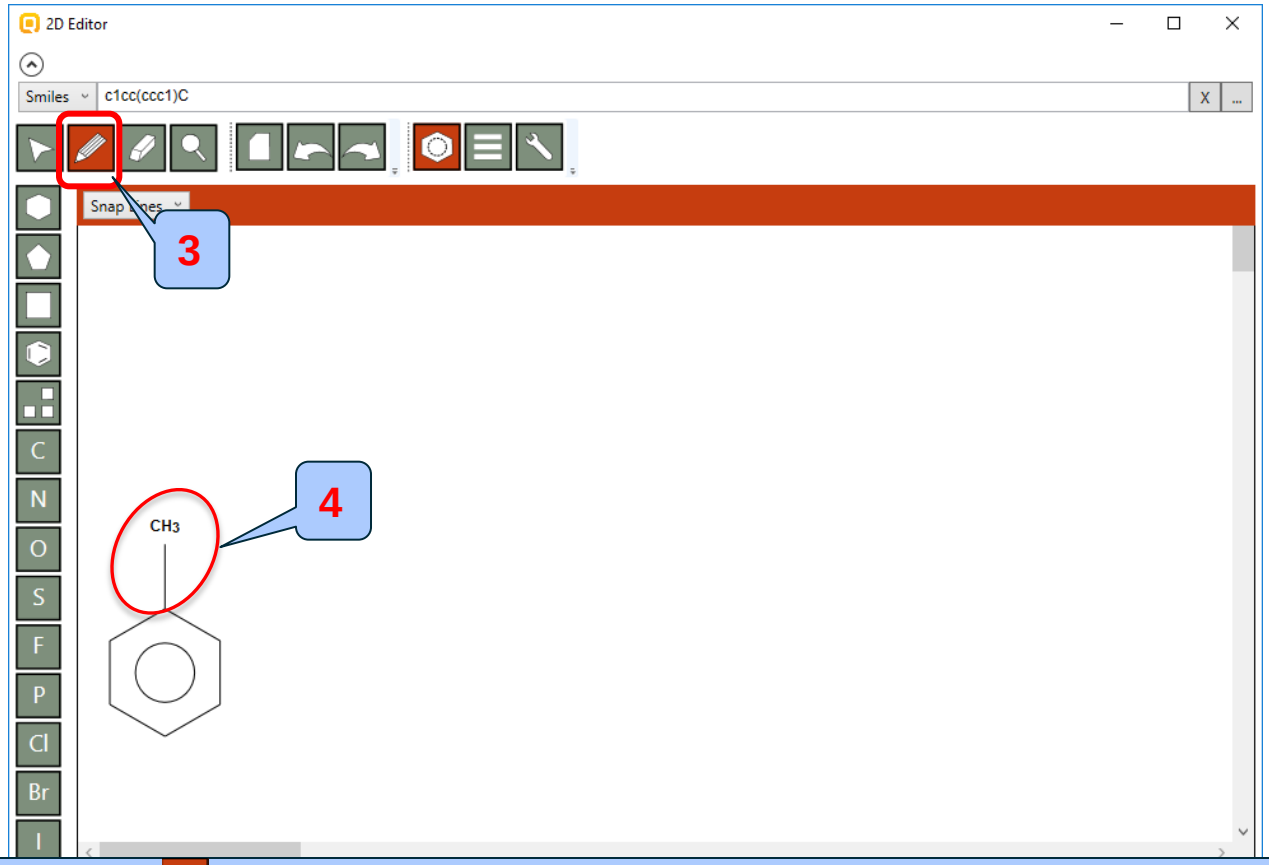
Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor



1. Left-click on the appropriate template form (in this case the benzene ring).
2. Move the cursor to the large clear area and left-click again, this puts the selected template in the plot.

Chemical Input

Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor



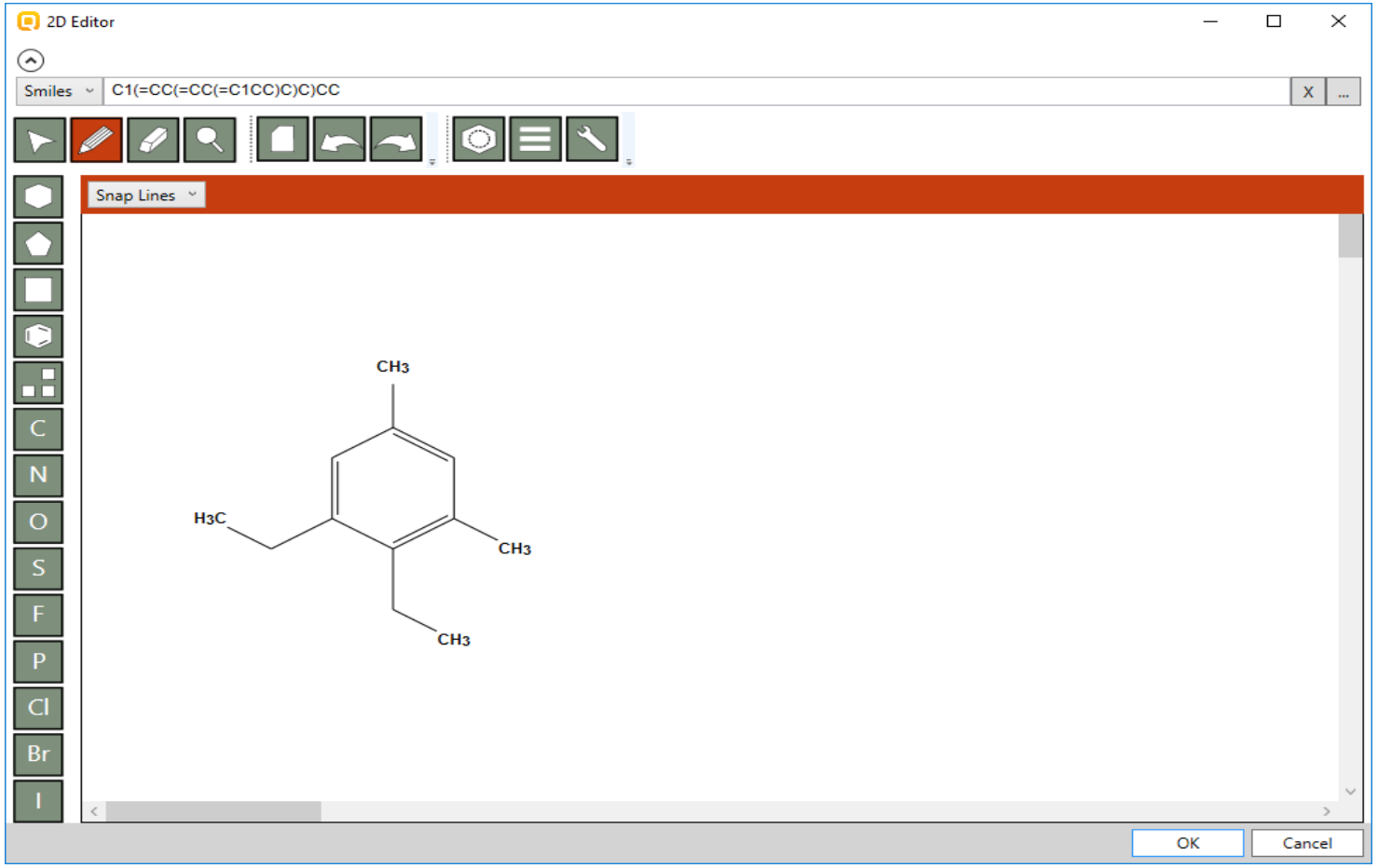
3. Click on the pencil button to draw a single bond.
4. Drag the mouse (pointing finger) to the appropriate atom and left-click to create a single bond.

Chemical Input by Drawing

- Note the default is addition of a CH_3-group.
- By moving the 'finger' to other C-atoms and left clicking the mouse adds other hydrocarbon fragments.
- If you make an incorrect entry you can click the 'undo' icon () in the upper corner of the screen to remove the addition.
- This process allows you to build the hydrocarbon skeleton of the target molecule (see next screen shot).

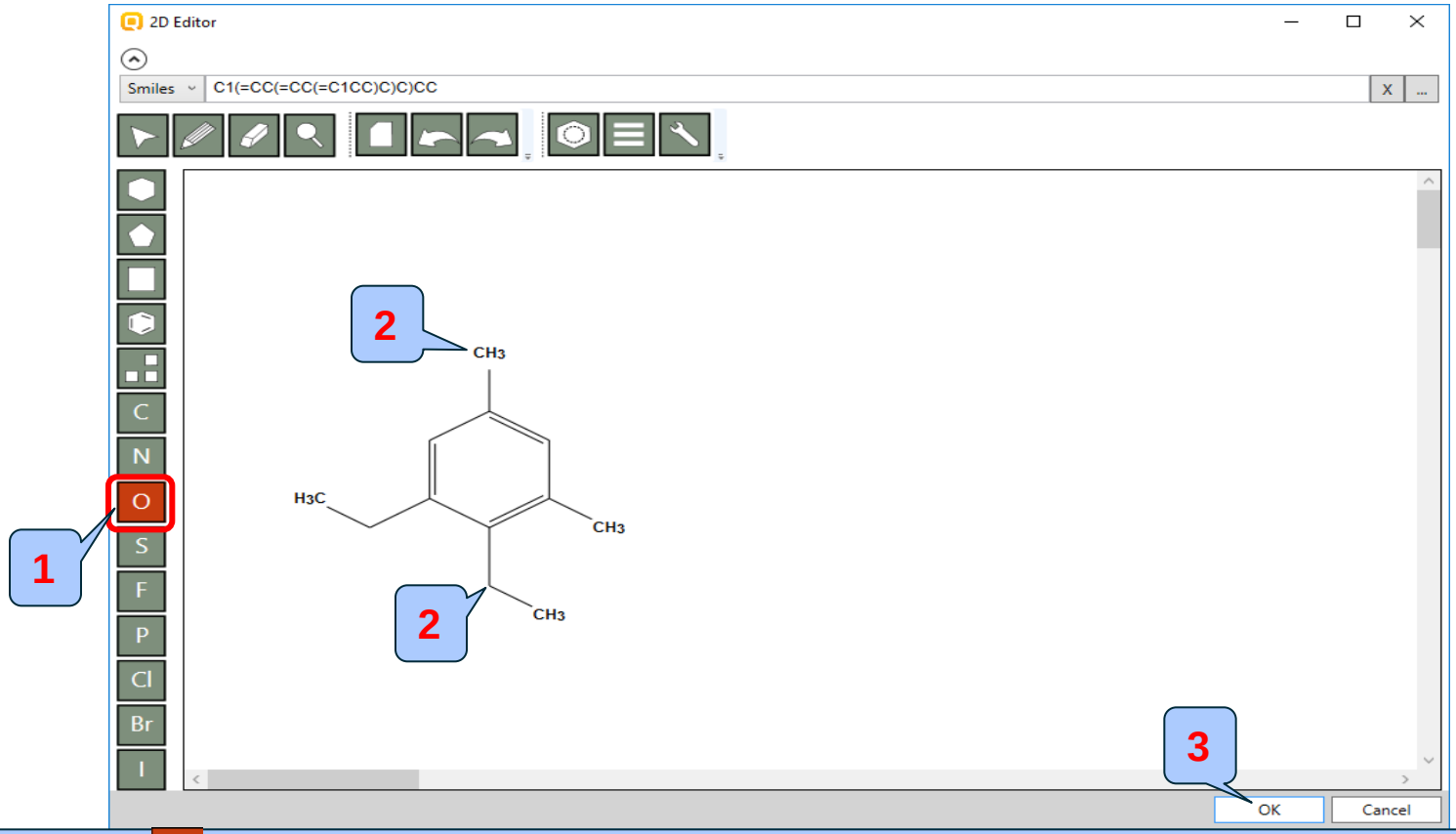
Chemical Input

Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor



Chemical Input

Drawing the target "3-ethyl-5-methyl-4-methoxyphenol" by 2-D editor

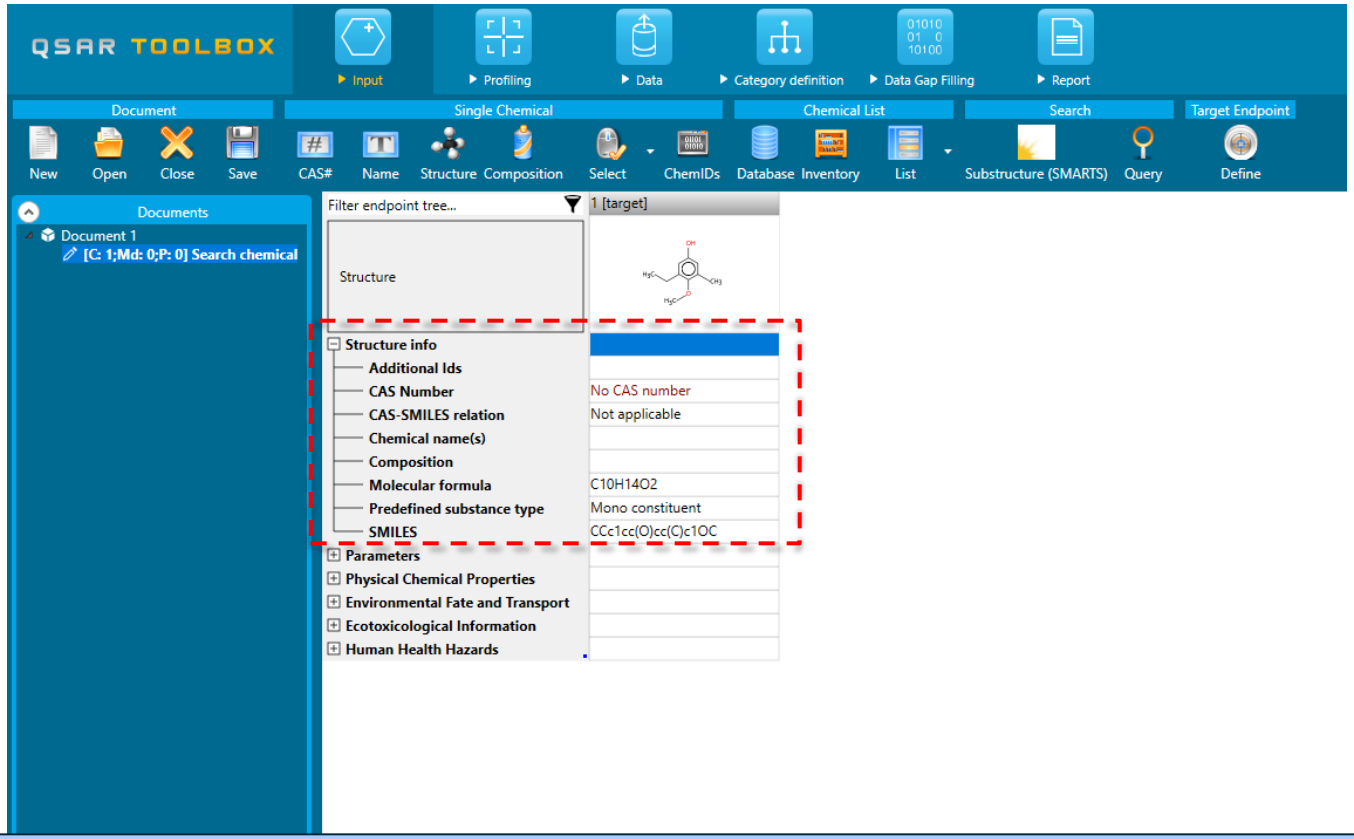


1. Click O to add an oxygen atom; 2. Left-click with mouse over the methyl group to insert an oxygen atom; 3. Click **OK**.

Chemical Input
Target chemical identity

- The already drawn target structure automatically appears on the data matrix
- Note that no CAS number or name is displayed for this chemical. This means the target chemical is not listed in the chemical inventories/databases implemented in the Toolbox (see next slide).

Chemical Input
Target chemical identity



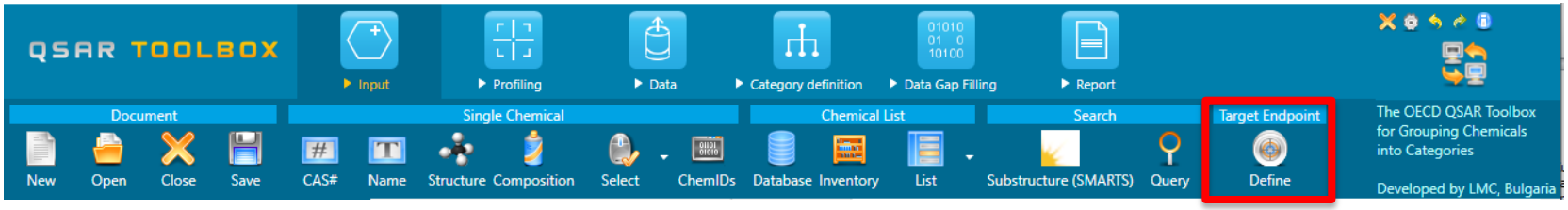
The workflow on the first module is now complete, and the user can proceed to the next module.

Outlook

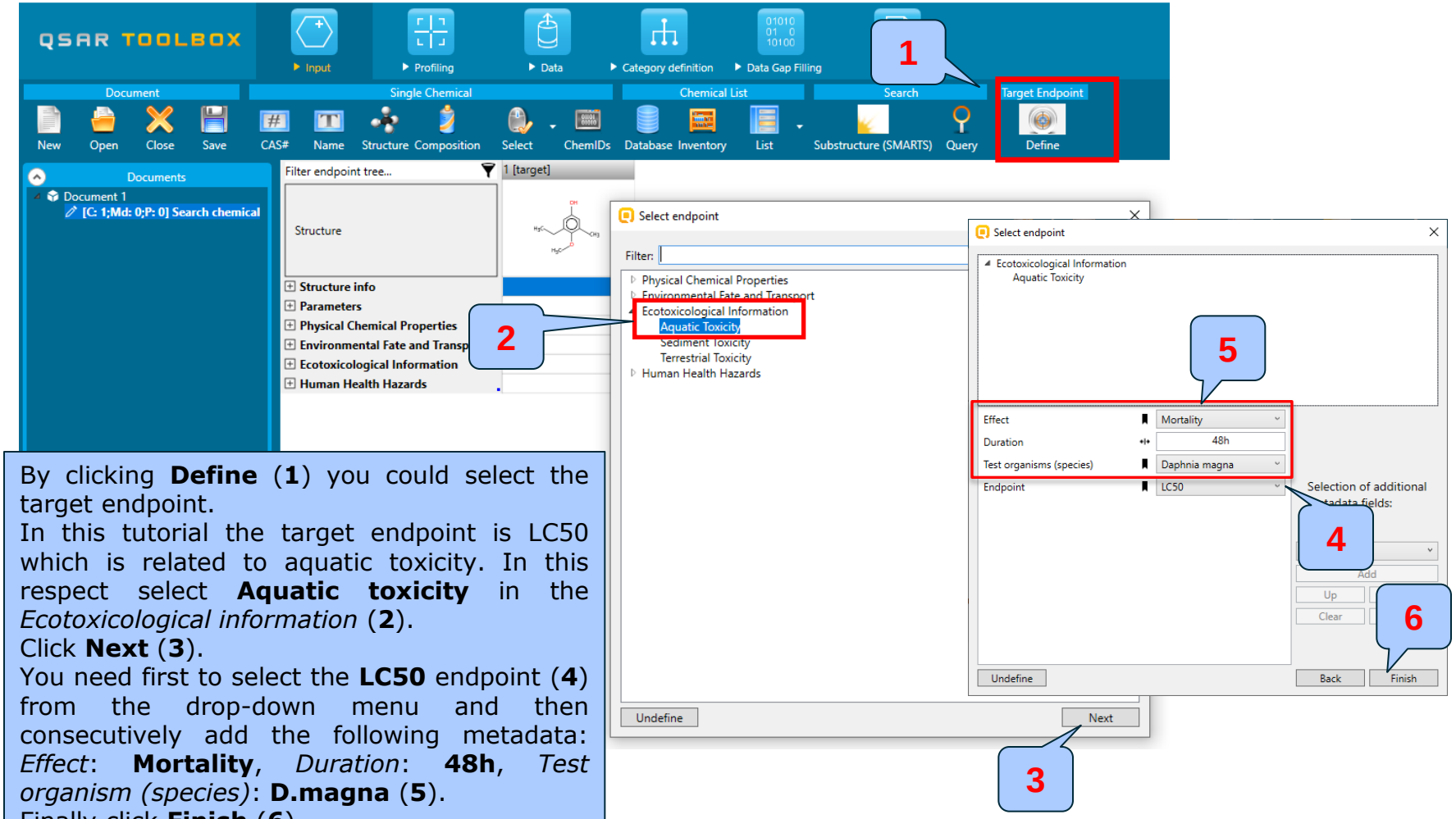
- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
    - **Define target endpoint**

Input
Define target endpoint

- Defining of the endpoint allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc., along with specific metadata information. Based on the metadata, relevancy of the profiles and databases is provided expressed in different color highlighting:
  - In green are highlighted the most suitable profilers related to the endpoint and databases including data for the defined target endpoint,
  - In orange are colored profilers which are plausible with respect to the defined target endpoint.

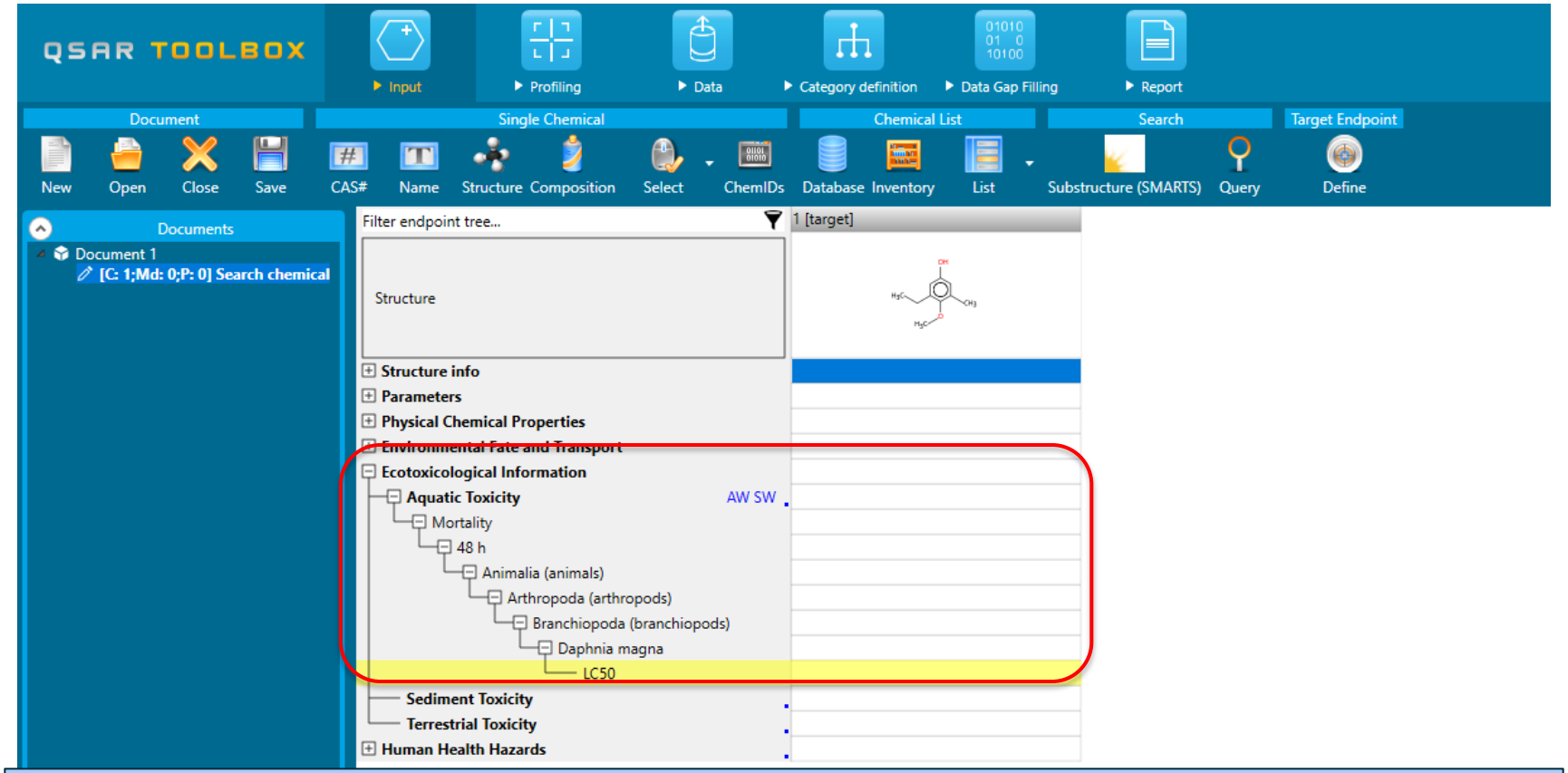


Input
Define target endpoint



By clicking **Define** (**1**) you could select the target endpoint.
In this tutorial the target endpoint is LC50 which is related to aquatic toxicity. In this respect select **Aquatic toxicity** in the *Ecotoxicological information* (**2**).
Click **Next** (**3**).
You need first to select the **LC50** endpoint (**4**) from the drop-down menu and then consecutively add the following metadata: *Effect*: **Mortality**, *Duration*: **48h**, *Test organism (species)*: **D.magna** (**5**).
Finally click **Finish** (**6**).

Input
Define target endpoint



The endpoint tree is automatically expanded to the level of the defined endpoint and the row is highlighted in yellow

Input
Input results

1) In module *Input*, you have entered the target chemical by drawing it.
2) The target endpoint (aquatic toxicity, LC50) is defined using "Define target endpoint" functionality.
3) Based on the defined target endpoint the relevant profiles and databases become highlighted in color (see next slides).

Outlook

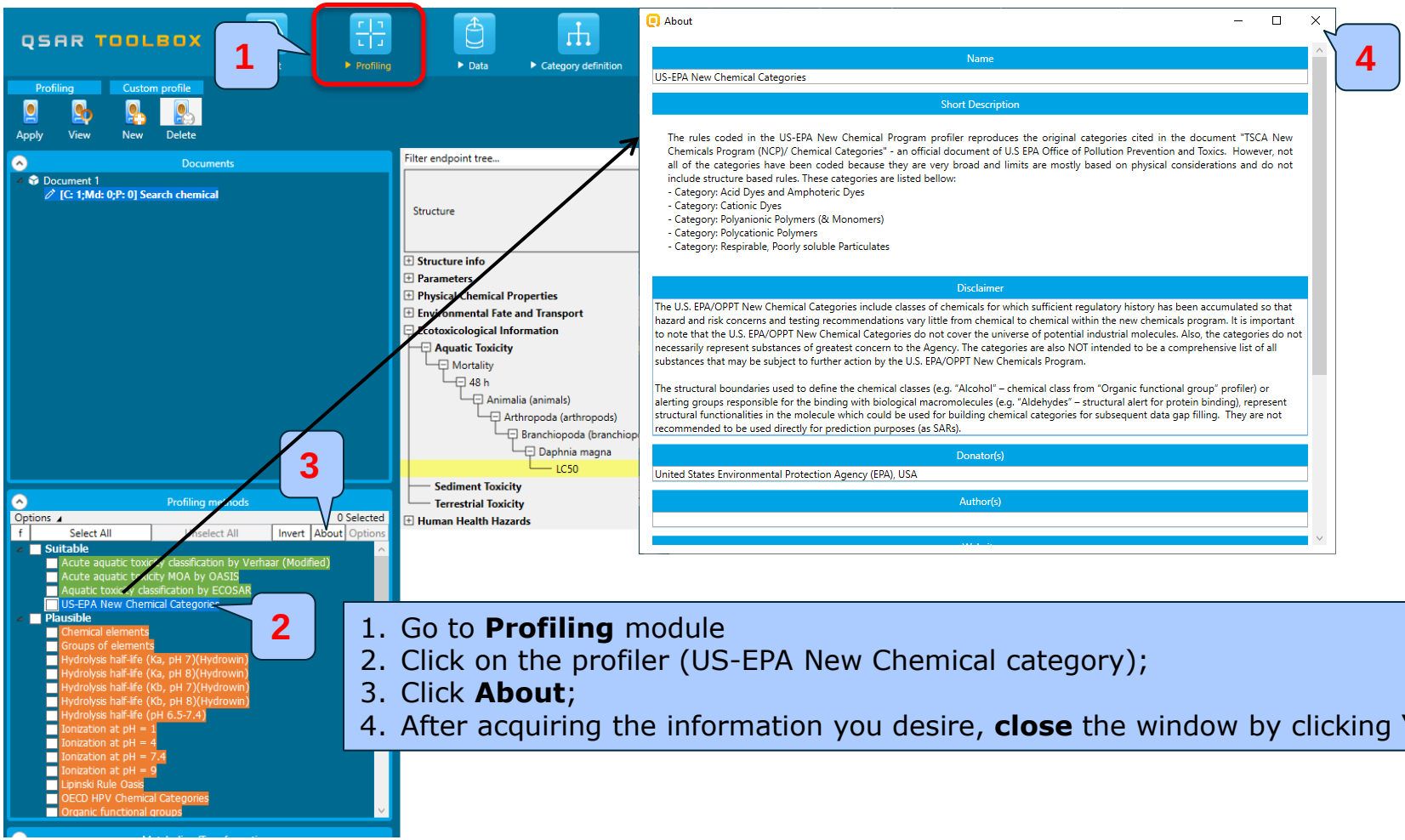
- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - **Profiling**

Profiling
Overview

- *"Profiling"* module refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database;
- Available "profilers" includes likely mechanism(s) of action, wich could be useful in forming categories that include the target chemical;
- "Profilers" are a collection of empirical and mechanism knowledge which could be used to analyse the structural properties of chemicals;
- The "profilers" identify the affiliation of the target chemical(s) to preliminary defined categories (functional groups/alerts);
- The "Profiling" module contains also observed and simulated metabolisms/transformations, which could be used in combination with the profilers;
- The outcome of the profiling determines the most appropriate way to search for analogues, but they are also useful for preliminary screening or prioritization of substances;
- The "profilers" are not (Q)SARs, i.e. they are not prediction models themselves;
- Based on the "profilers' relevancy" (determined by the defined target endpoint), the most suitable and plausible once are getting colour highlighted.

Profilers
Background

Summary information of the different profilers are provided in the “About”.

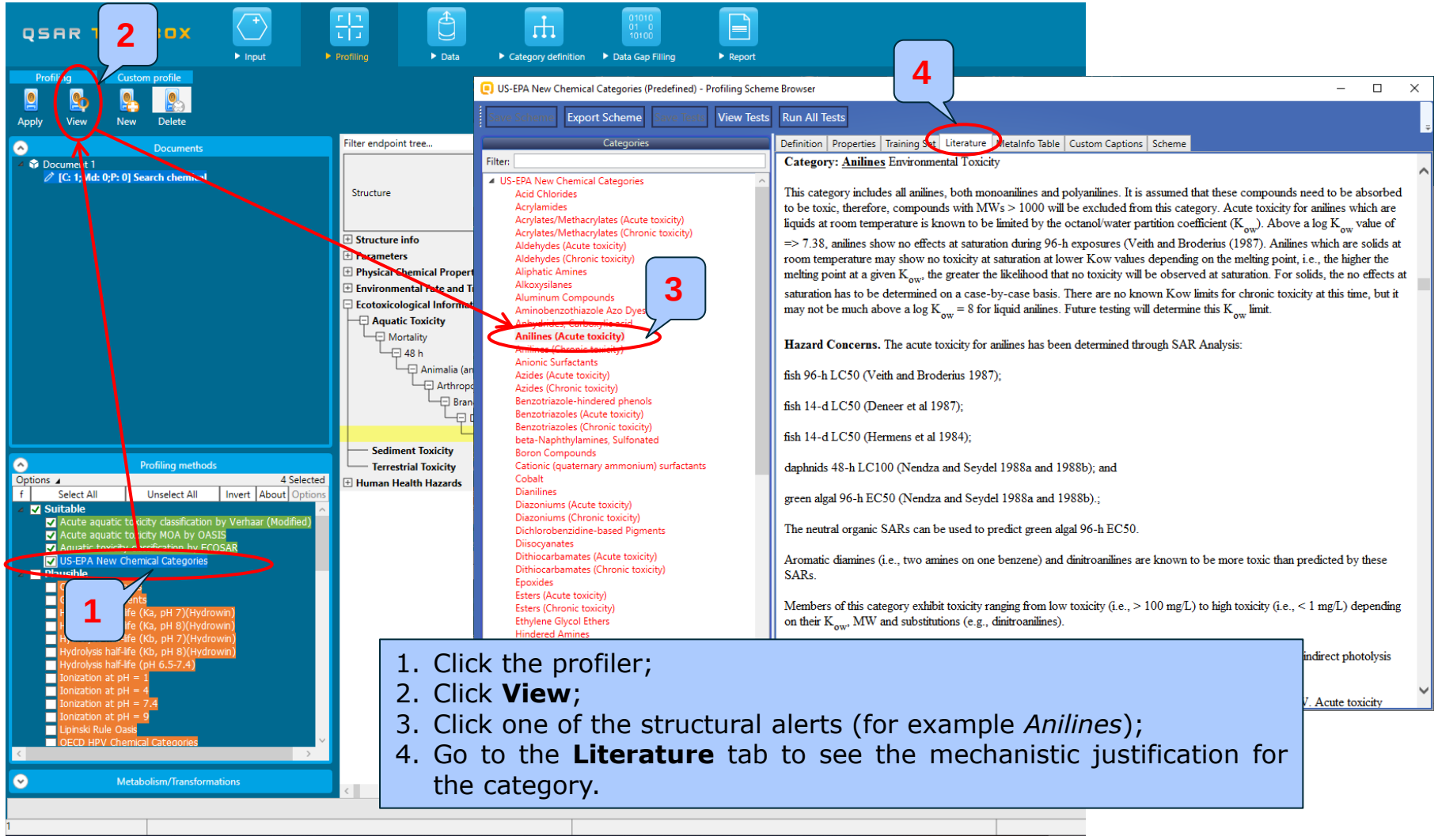


1. Go to **Profiling** module
2. Click on the profiler (US-EPA New Chemical category);
3. Click **About**;
4. After acquiring the information you desire, **close** the window by clicking “**x**”;

Profilers
Background

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, *US-EPA New Chemical Categories*) and clicking **View** (see next slide).

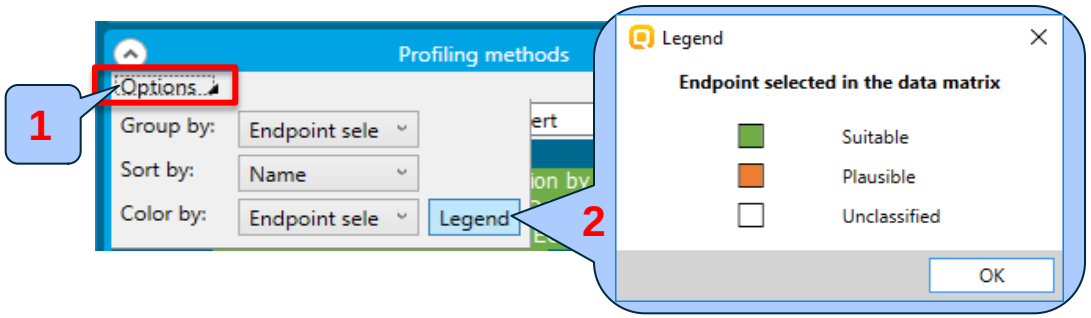
Profilers
Background



1. Click the profiler;
2. Click **View**;
3. Click one of the structural alerts (for example *Anilines*);
4. Go to the **Literature** tab to see the mechanistic justification for the category.

Profilers
Background

- Once the endpoint is selected, the relevant profiles and metabolic transformations are highlighted. The meaning of the colors can be seen within the **Options** (1) by click **Legend** (2).

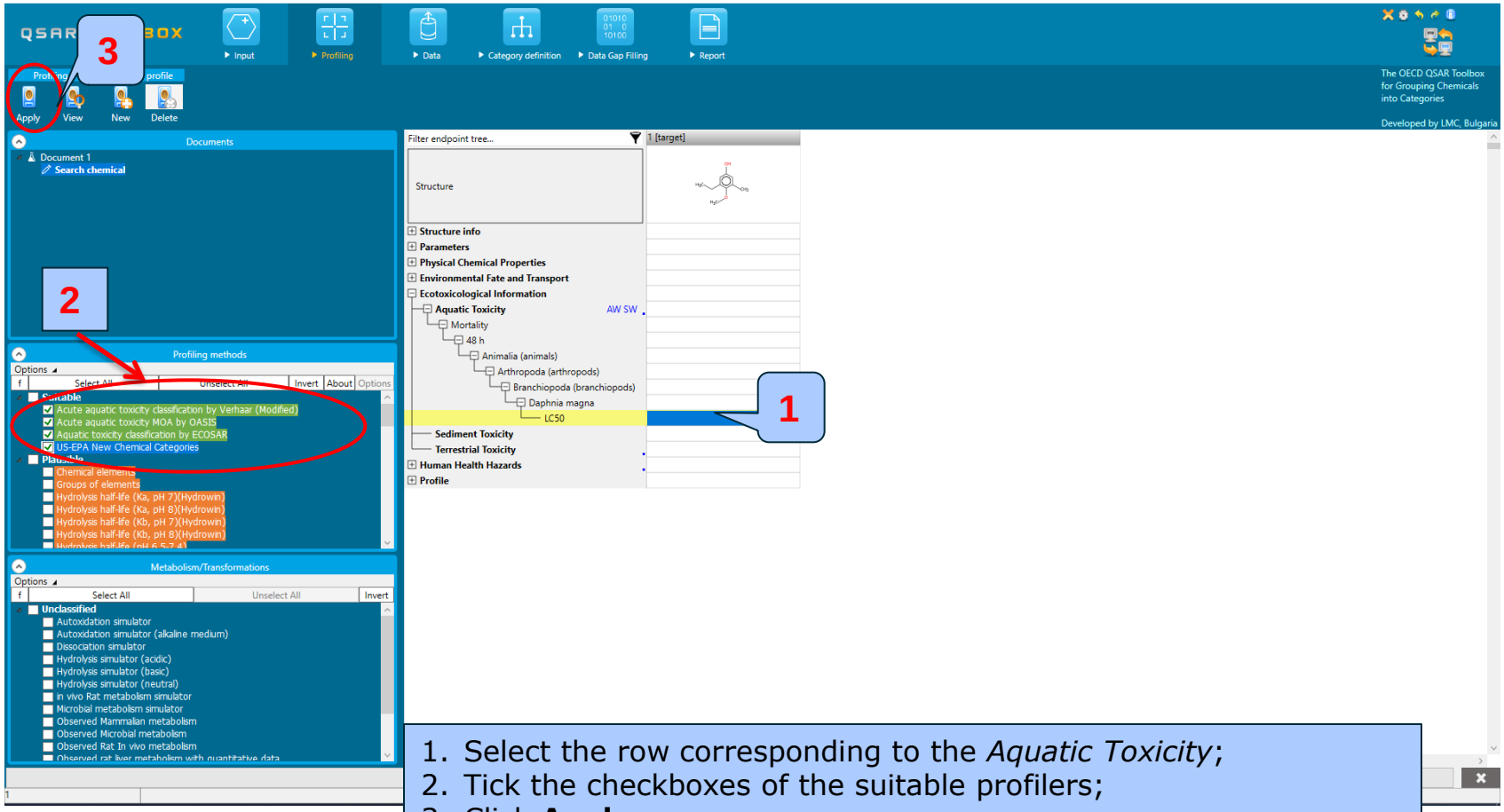


- **Suitable** - developed using data/knowledge for the target endpoint;
- **Plausible** – not endpoint specific; structure-based; form broader group of analogues;
- **Unclassified** – all profilers, which are not classified in any of the categories above.

- Select the Profiling methods related to the target endpoint by ticking the checkbox next to the profilers name.
- This selects (a green check mark appears) or deselects (the green check disappears) profilers.

Profiling

Profiling the target chemical



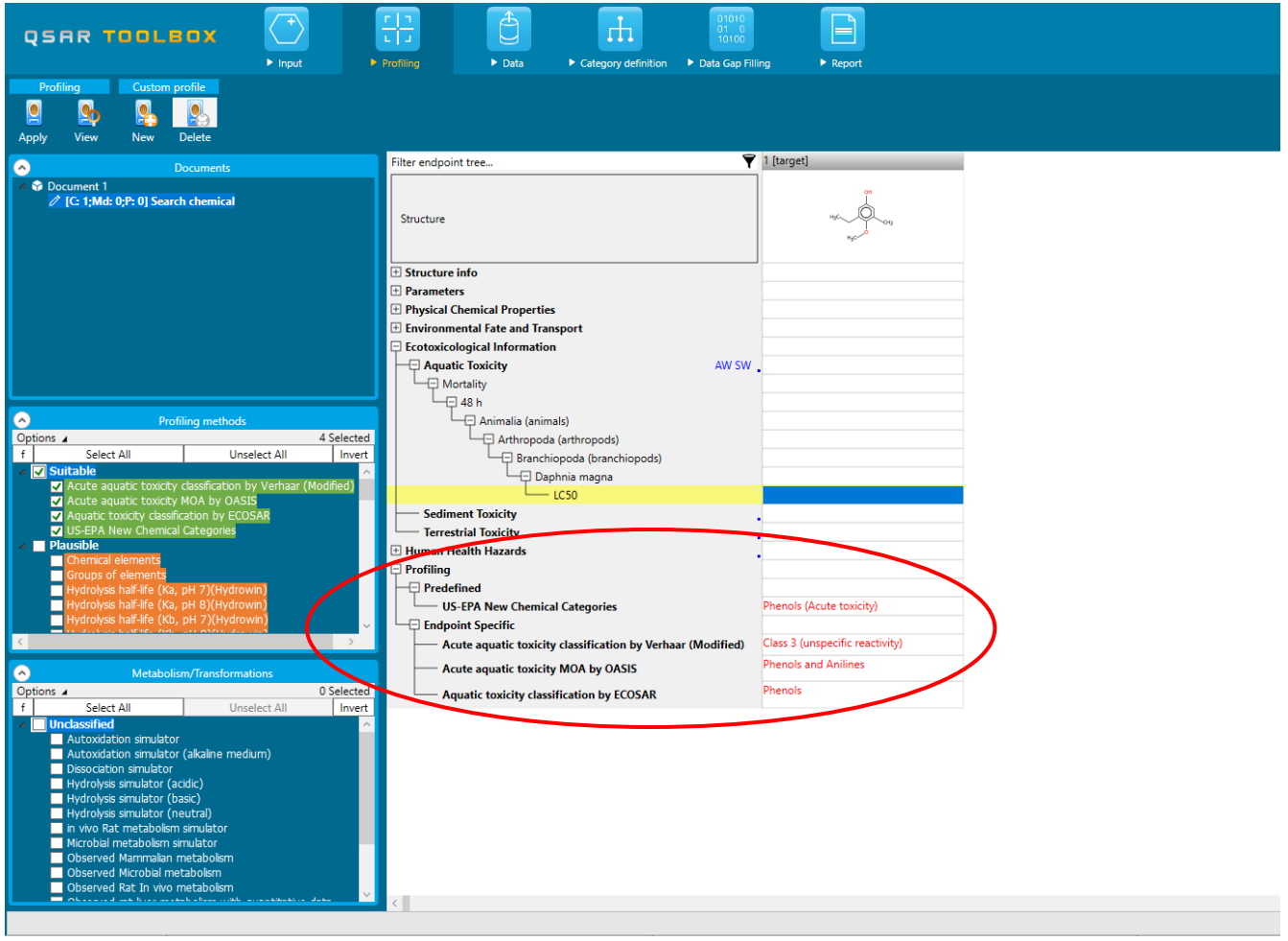
1. Select the row corresponding to the *Aquatic Toxicity*;
2. Tick the checkboxes of the suitable profilers;
3. Click **Apply**.

Profiling
Profiling the target chemical

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appear as a dropdown box under the target chemical.
- Please note the specific profiling results obtained by the most suitable profilers.
- These results will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiles of the target "3-ethyl-5-methyl-4-methoxyphenol



Profiling
Profiling results

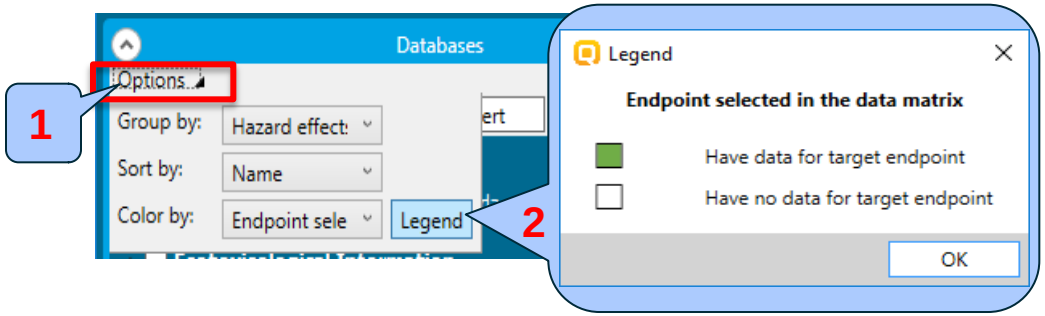
1) In module *Profile*, you have profiled the target chemical according to the suitable profilers (green) related to the target endpoint.
2) The target chemical is categorized as "phenol" based on predefined Acute aquatic toxicity US-EPA profiler (hereafter called US-EPA) and the two endpoint-specific profilers (Acute aquatic toxicity classification by ECOSAR (hereafter called ECOSAR) and Acute aquatic toxicity MOA by OASIS (hereafter called MOA)
3) By the endpoint-specific "Acute aquatic toxicity classification by Verhaar" the target is categorized as "Class 3 (unspecific reactivity)"
4) In general the target is classified as "phenol"
5) All of the above mentioned profilers could be used for categorization purposes (collecting analogues)
6) In this case ECOSAR profiler will be used for categorization purpose (primary grouping).

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - **Data**

Data

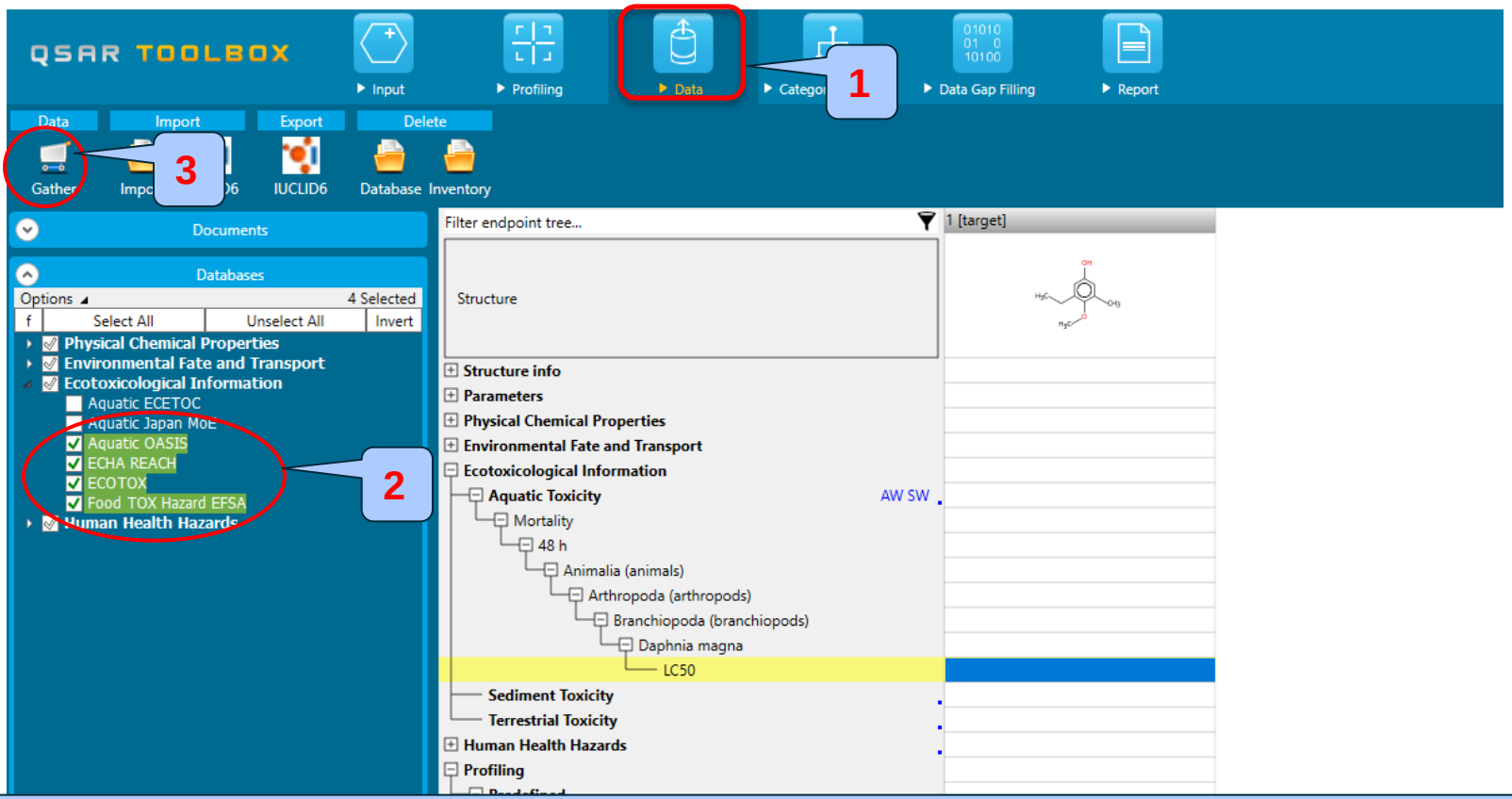
- *Data* module refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox databases.
- Data gathering can be executed in a global fashion (i.e. collecting all data of all endpoints) or on a more narrowly defined basis (i.e. collecting data for a single or limited number of endpoints).
- Once the endpoint is selected, the relevant databases are highlighted. Meaning of the colors could be seen within the **Options** (1) by click **Legend** (2).



- In this example, we limit our data gathering to the common aquatic toxicity endpoints from databases containing aquatic toxicity data (Aquatic OASIS, ECHA REACH, ECOTOX, Food TOX Hazard EFSA).

Data

Gather data using database relevancy

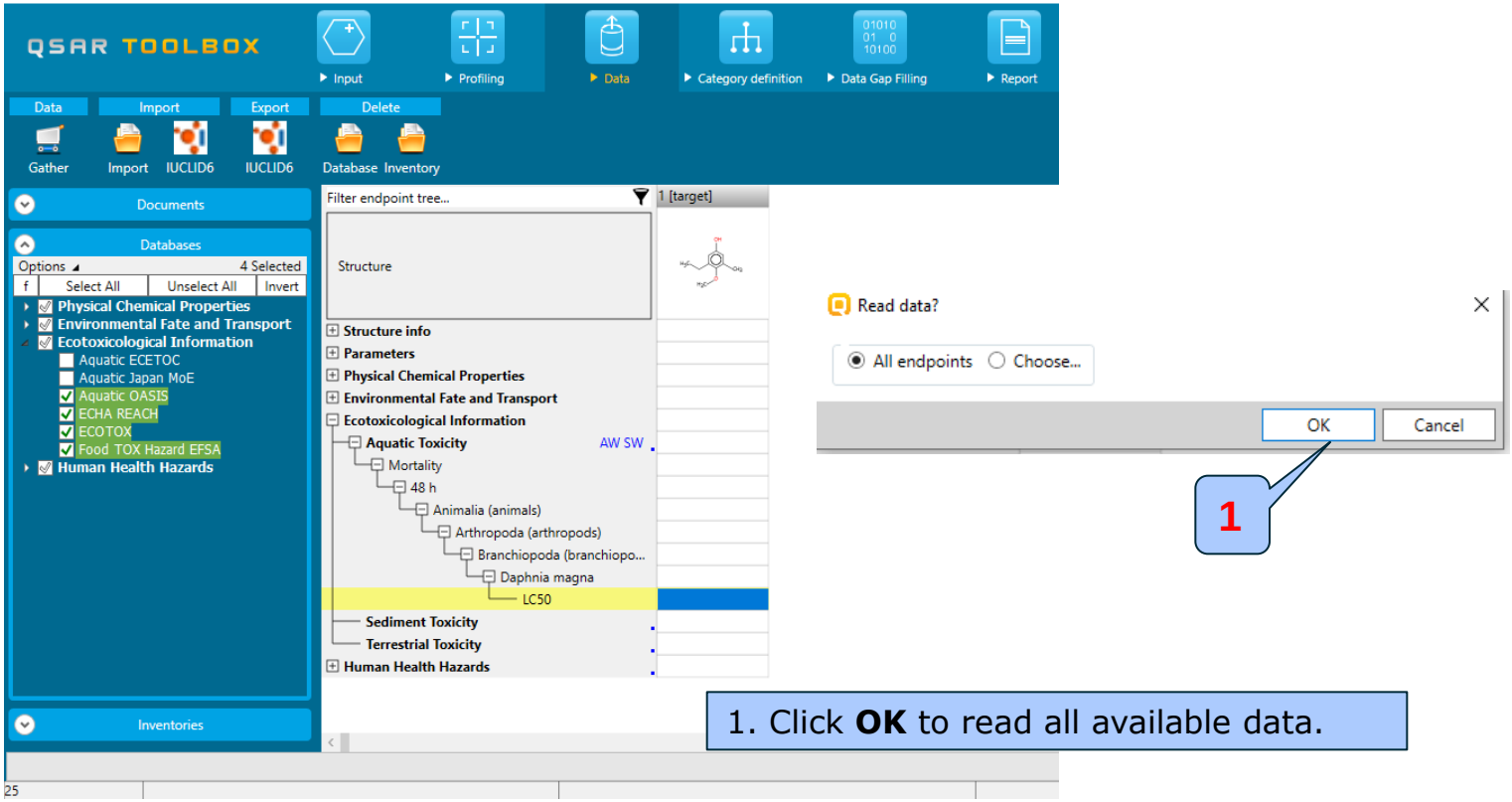


1. Go to **Data** module;
2. Select the green highlighted databases corresponding to the *Aquatic toxicity (Aquatic OASIS, ECHA REACH, ECOTOX, Food TOX Hazard EFSA);*
3. Click **Gather**.

Data
Process of collecting data

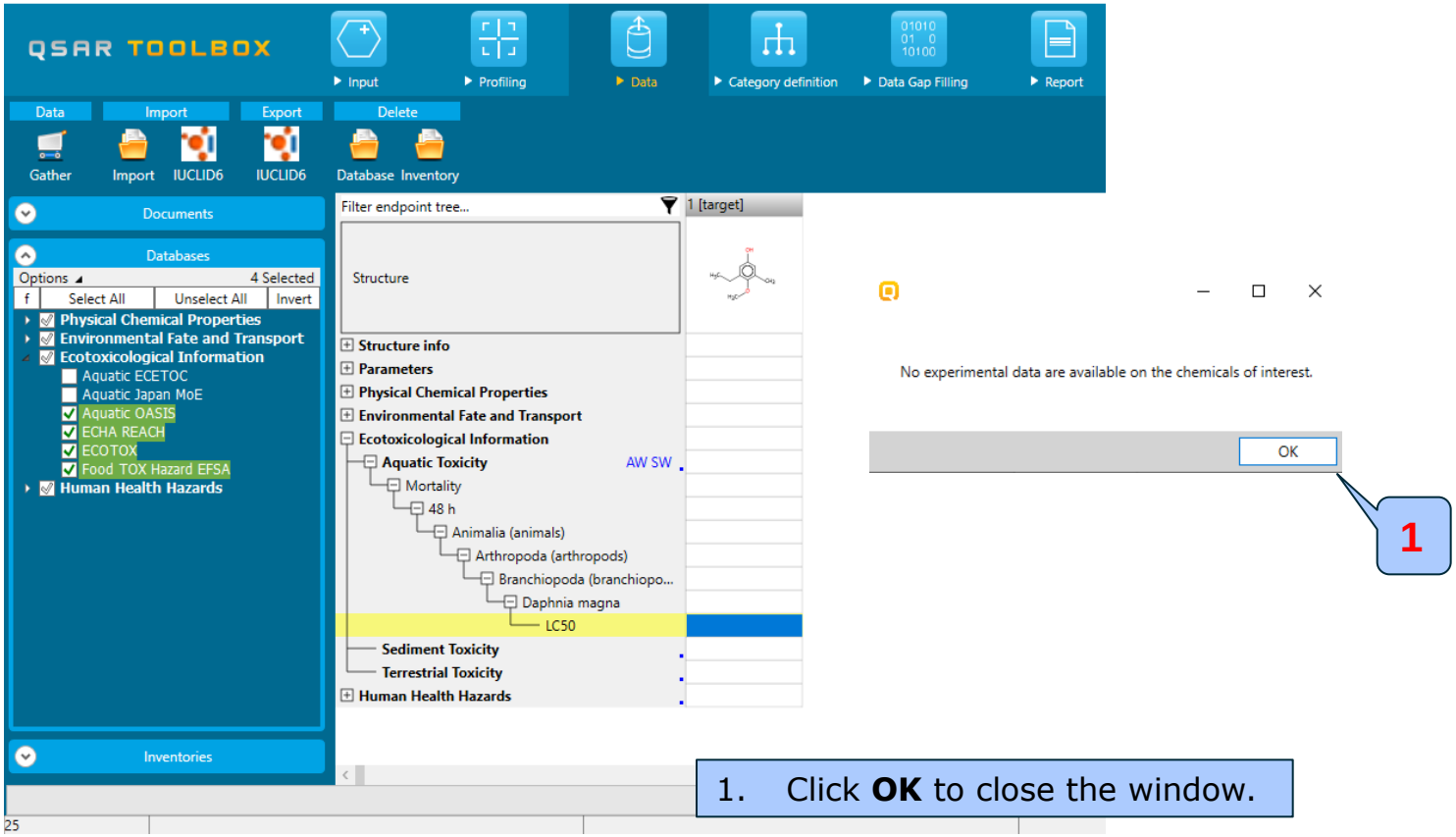
Toxicity information on the target chemical is electronically collected from the selected datasets.

A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.



Data
Process of collecting data

In this example, an insert window appears stating that no experimental data is available for the chemical of interest.



Recap

- You have entered the target chemical being sure of the correct structure.
- You have profiled the target chemical and found no experimental data is currently available for this structure.
- In other words, you have identified a data gap, which you would like to fill.
- Now you are ready to continue with next step of the workflow - *Category Definition*.

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - Data
  - **Category definition**

Category Definition
Overview

- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in defining the category definition.

Category Definition
Grouping methods

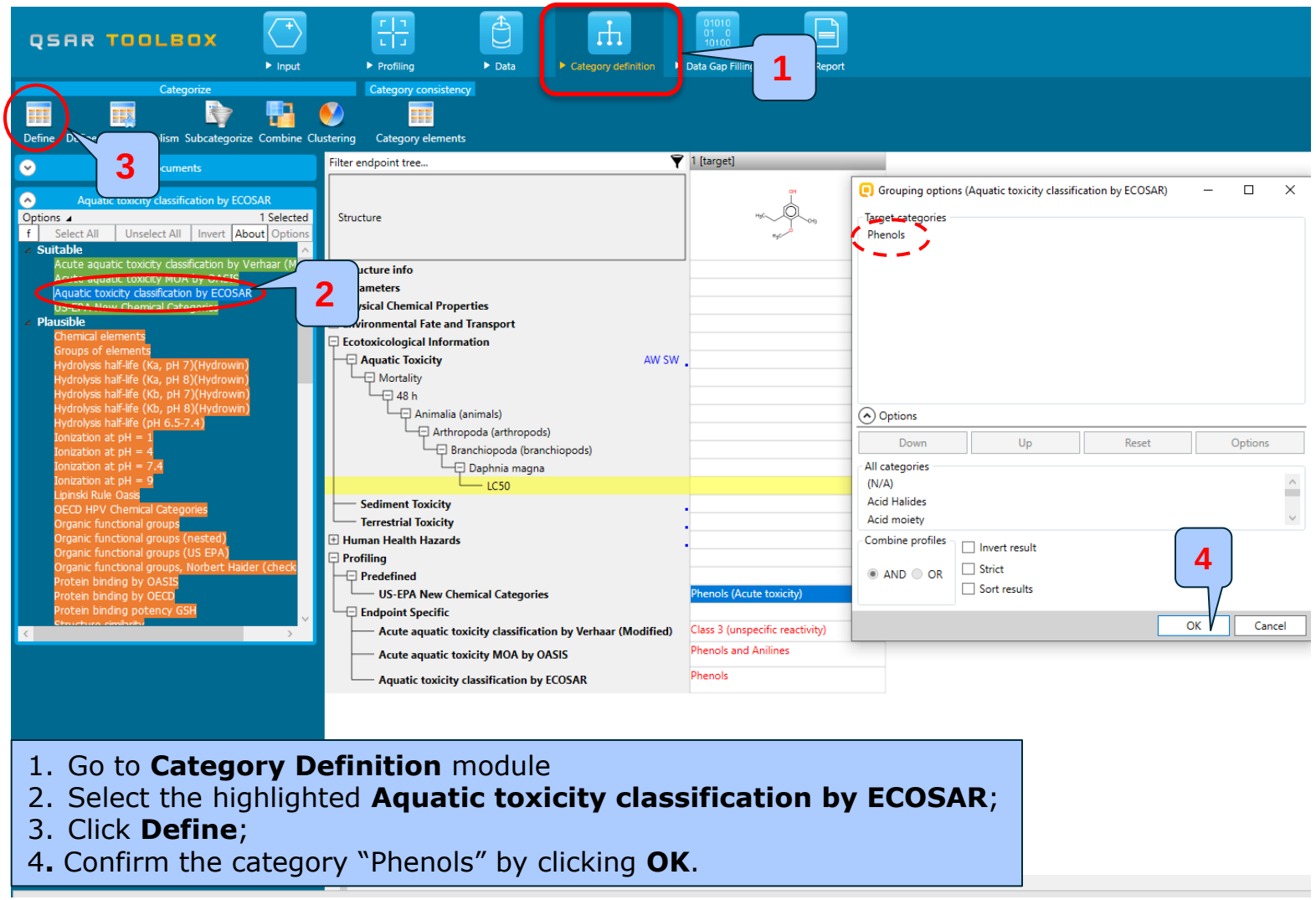
- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of "similarity" so that within a category data gaps can be filled by trend-analysis.
- For this example, starting from the target chemical a specific ECOSAR classification is identified ("phenols"), subsequently analogues are found within the same specific classification ("phenols") for which experimental results are available.

Category Definition
ECOSAR categories - overview

- ECOSAR has been used by the U.S. Environmental Protection Agency since 1981 to predict the aquatic toxicity of new industrial chemicals in the absence of test data.
- The *Aquatic toxicity classification by ECOSAR* profiling scheme in the Toolbox is used for grouping of chemicals by structural similarity which may or may not have mechanistic meaning. Experience has shown ECOSAR to be a robust profiler which makes it a logical choice in an initial profiling scheme.

Category Definition

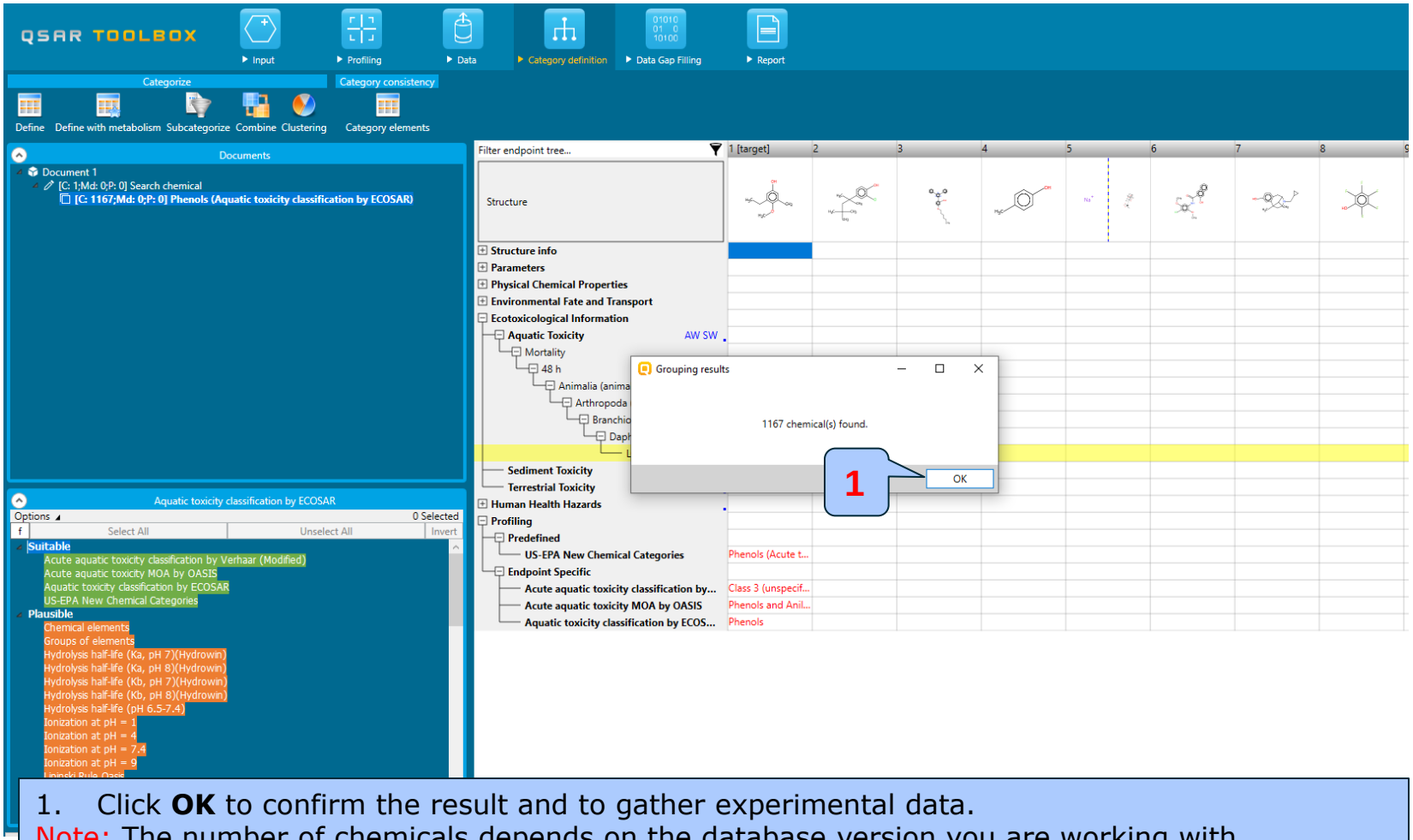
Defining ECOSAR category



1. Go to **Category Definition** module
2. Select the highlighted **Aquatic toxicity classification by ECOSAR**;
3. Click **Define**;
4. Confirm the category "Phenols" by clicking **OK**.

Category Definition

Defining ECOSAR category



1. Click **OK** to confirm the result and to gather experimental data.

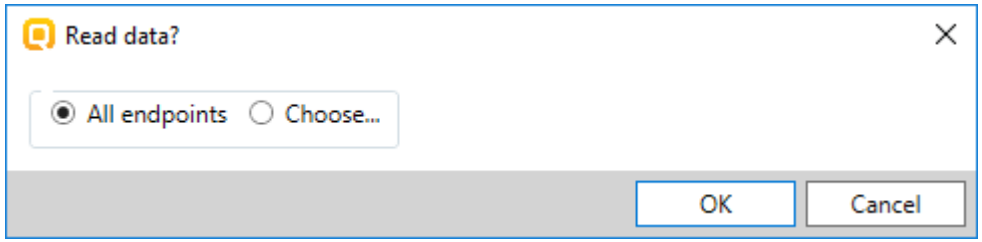
Note: The number of chemicals depends on the database version you are working with.

Category Definition
Analogues

- The Toolbox now identifies all chemicals corresponding to the ECOSAR classification of *Phenols* which are listed in the selected databases within the *Data* module.
- 1167 analogues are identified. Along with the target they form a category (Phenols) which can be used for data gap filling.

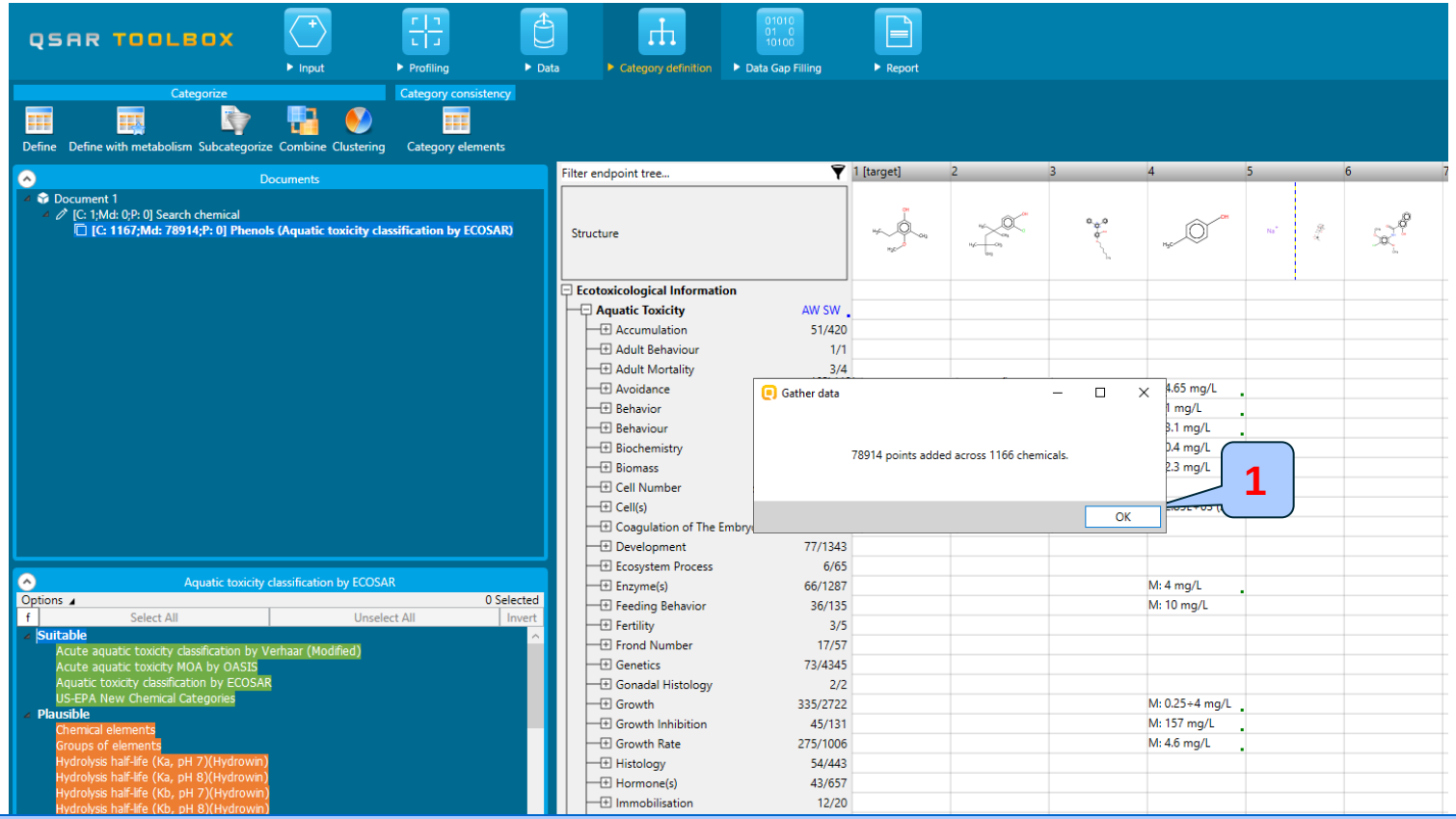
Category Definition
Read data for Analogues

- The Toolbox automatically requests the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



- In this example, since only databases that contain information for Eco-toxicological endpoints are selected, both options give the same results.
- As the Toolbox must search the database, this may take some time.

Category Definition
Summary information of Analogues

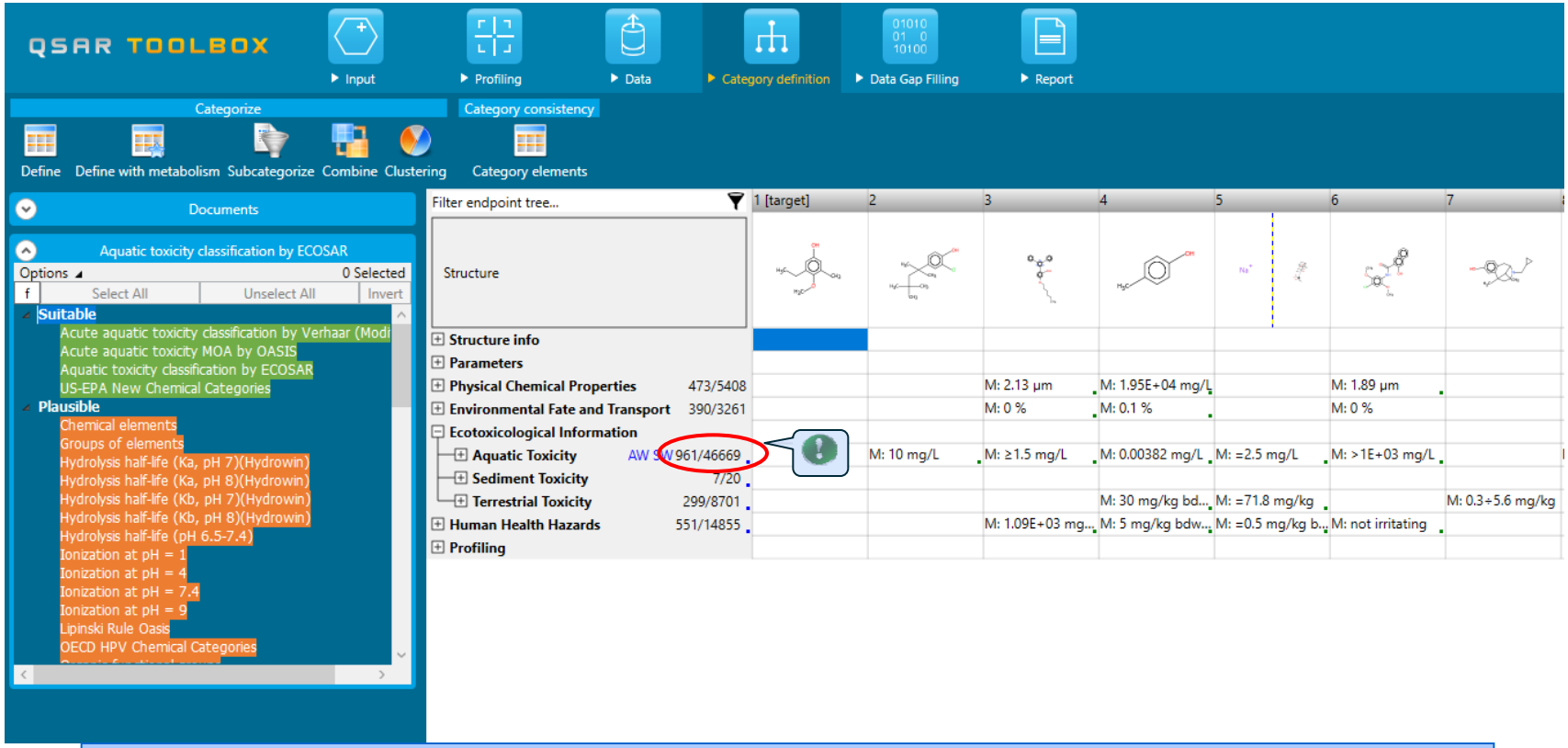


1. Click **OK** on window that provides information for common number collected data across the initial group of chemicals.

Note: The number of data points depends on the database version you are working with.

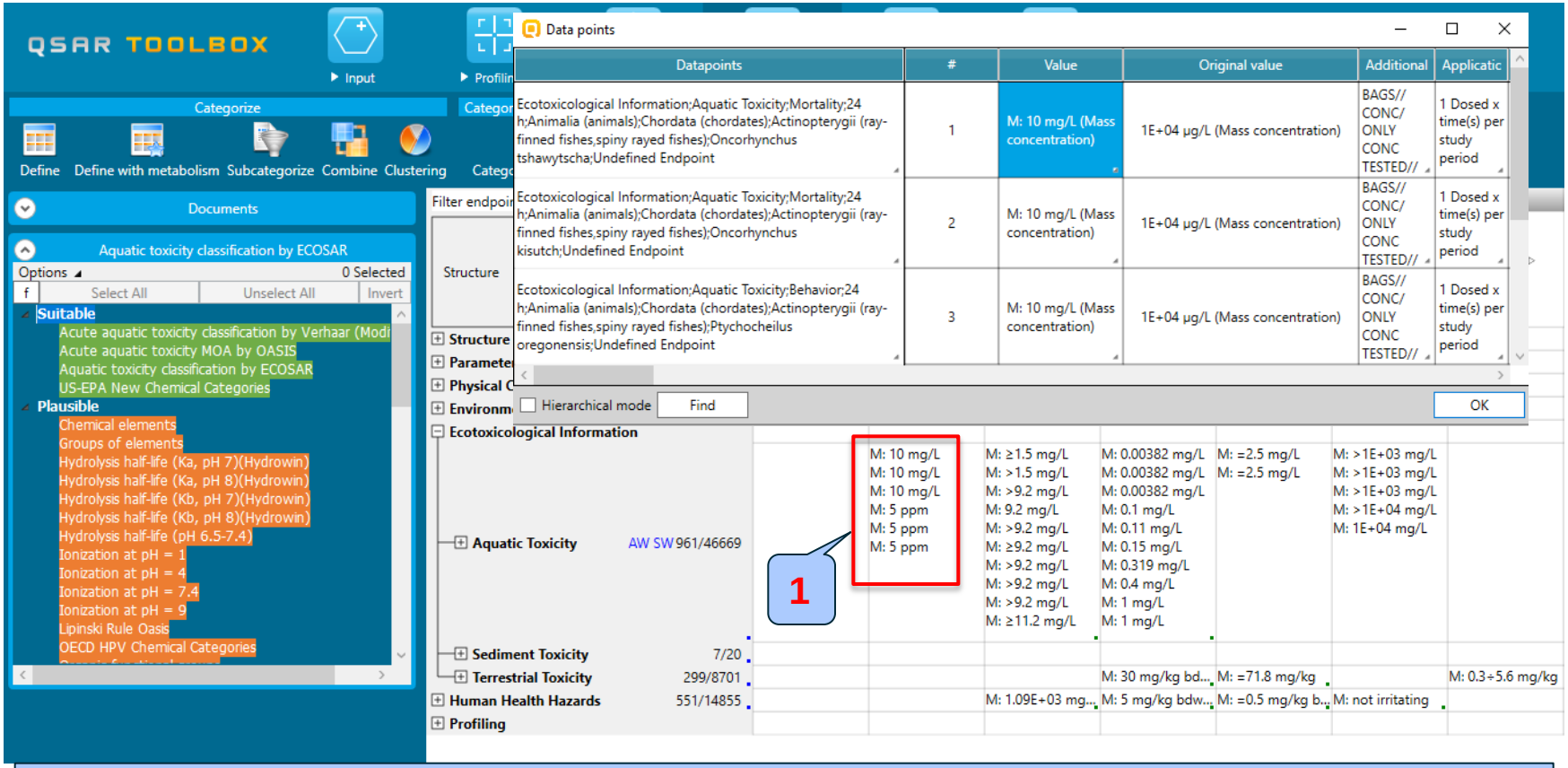
Category Definition

Summary information of Analogues



Chemical statistics presenting the number of chemicals and the available experimental data for them. This is statistics for the current row on data matrix.

Category Definition
Summary information of Analogues



1. Double-click on the cell with measured data opens a table which provides detailed information for all experimental data of the focused chemical.

Recap

- You have identified a category of 1166 chemicals (*Phenols*) with the *Aquatic toxicity classification by ECOSAR* profiler for the target chemical 3-ethyl-5-methyl-4-methoxyphenol.
- The available experimental results for these 1166 analogues have been collected from the selected databases (*Aquatic OASIS, ECHA REACH, ECOTOX and Food TOX Hazard EFSA*).
- But before the user can proceed with the *Filling Data Gap* module, he/she should navigate through the endpoint tree and find the specific gap that will be filled.

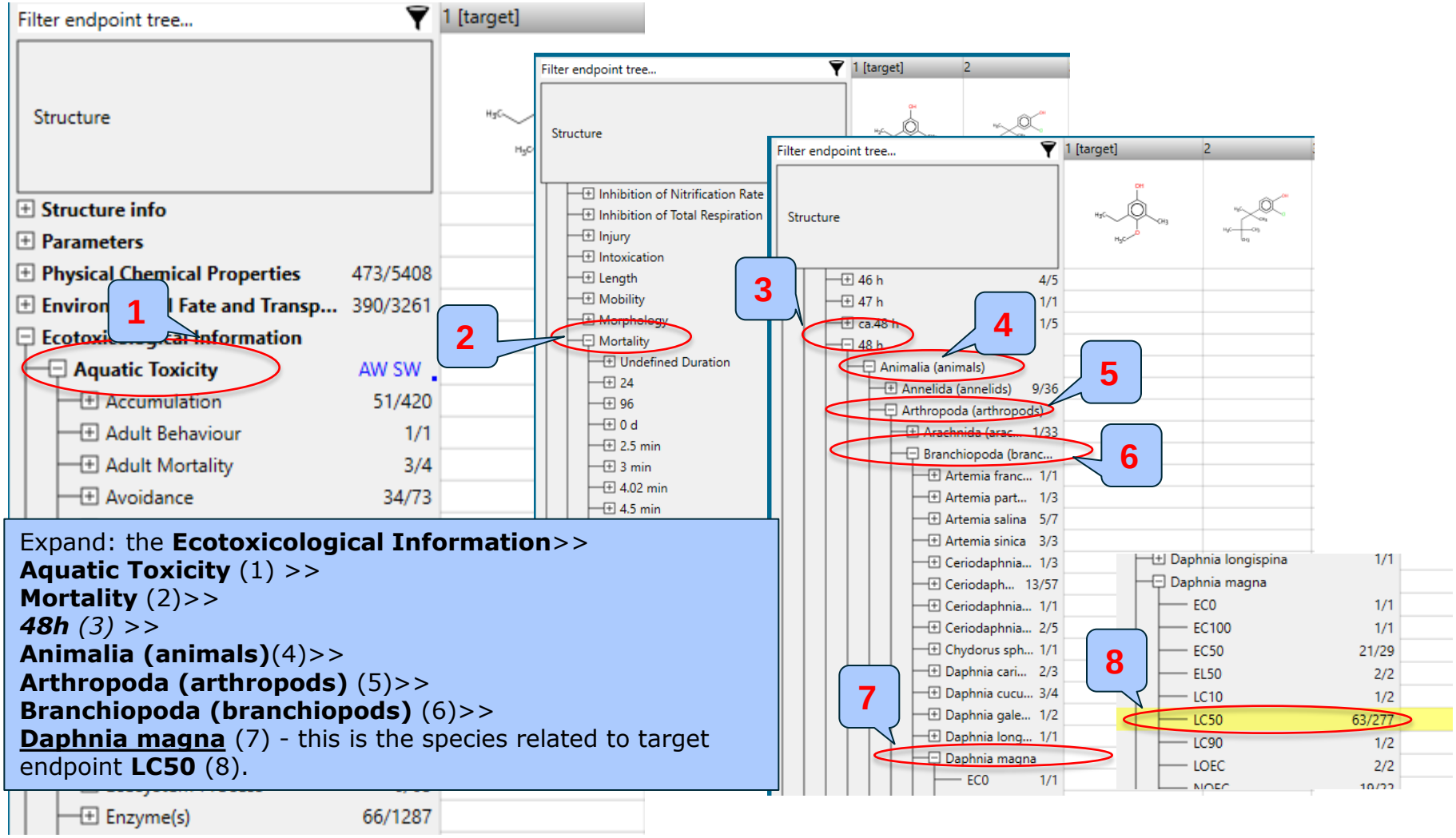
Category Definition

Navigation through the endpoint tree

- The user can navigate through the data tree by opening (or closing) the nodes of the tree.
- The data tree is extensive but logically constructed. It can be mastered with a practice.
- In this example, the "48 h LC50 Mortality for *Daphnia magna*" is the target endpoint.
- As already mentioned the row with the defined target endpoint is highlighted in yellow. (see next slide)

Category Definition

Navigation through the endpoint tree



Expand: the **Ecotoxicological Information**>>
Aquatic Toxicity (1) >>
Mortality (2)>>
48h *(3)* >>
Animalia (animals)(4)>>
Arthropoda (arthropods) (5)>>
Branchiopoda (branchiopods) (6)>>
Daphnia magna (7) - this is the species related to target endpoint **LC50** (8).

Recap

- You have now retrieved the available experimental data on aquatic toxicity for 1166 chemicals classified as "phenols" by the "ECOSAR" profiler found in the databases Aquatic OASIS, ECHA REACH, ECOTOX and Food TOX Hazard EFSA.
- You have identified the target endpoint of "48 h LC50 Mortality for *Daphnia magna*".
- You are ready to fill in the data gap so click *Data Gap Filling* (see next few slides).

Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - Data
  - Category definition
  - **Data Gap Filling**

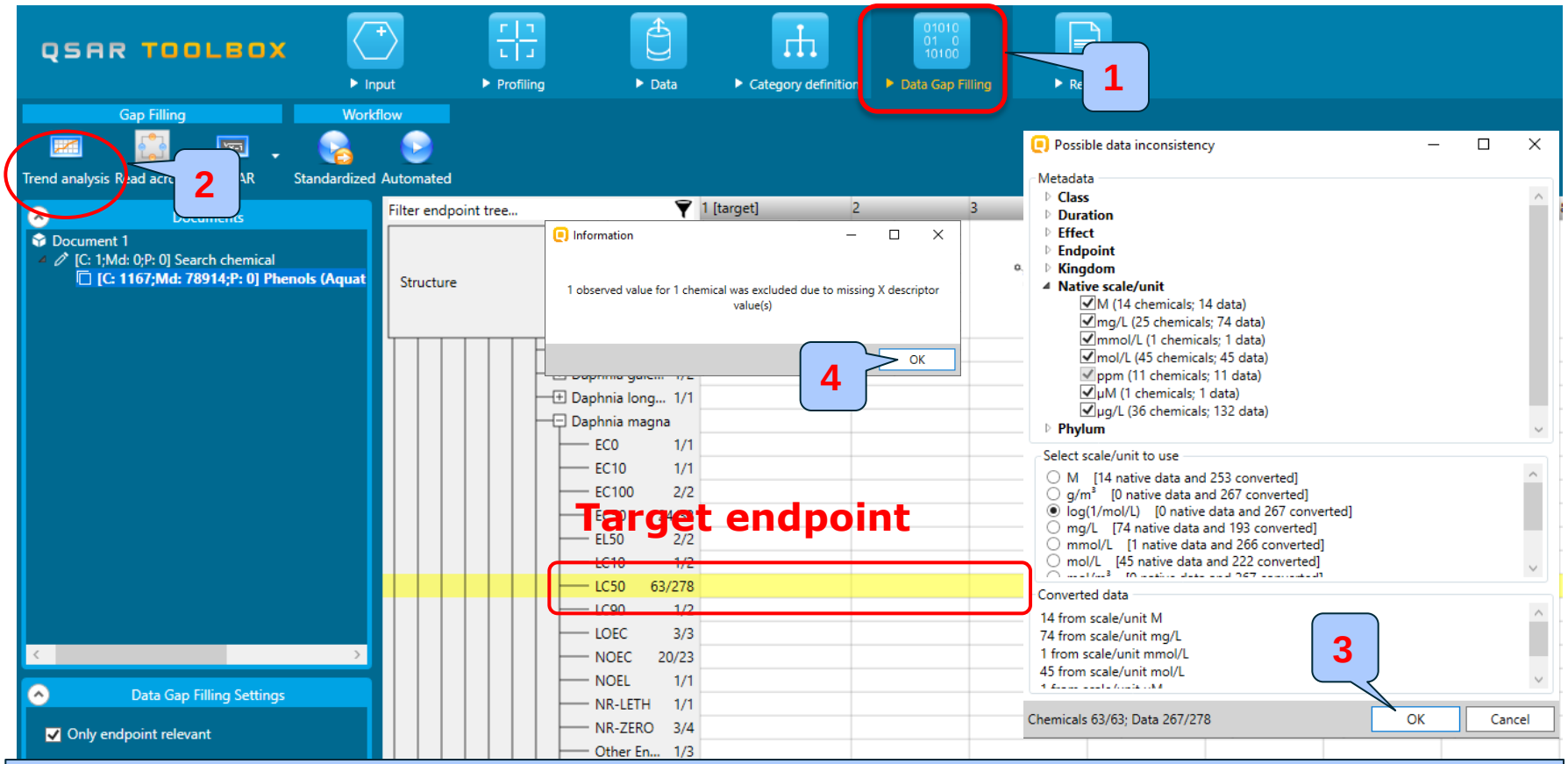
Data Gap Filling

Overview

- *Data Gap Filling* module gives access to five different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
  - Standardized workflow (SW)
  - Automated workflow (AW)
- The most relevant data gap mechanism is used , taking into account the following considerations:
  - *Read-across* is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
  - *Trend analysis* is the appropriate data-gap filling method for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
  - *(Q)SAR models* can be used to fill a data gap if no adequate analogues are found for a target chemical.
  - *Automated and standardized workflows* follow preliminary implemented logic. The AW is not affected by the user activities (proceeding or subsequent), while the SW stops at the each step of the workflows allowing the user to make different selection.
- In this example we will use trend analysis.

Data Gap Filling

Apply Trend analysis



1. Go to ***Data Gap filling*** module; 2. Click **Trend analysis**; 3. A pop-up window alerting you to possible data inconsistencies appears. Click **OK**; 4. A pop-up message informs about the number of chemicals (e.g. mixtures or UVCB substances) that will not be included in the Trend analysis prediction due to missing X descriptor value(s), which by default is Log*Kow*. Click **OK**.

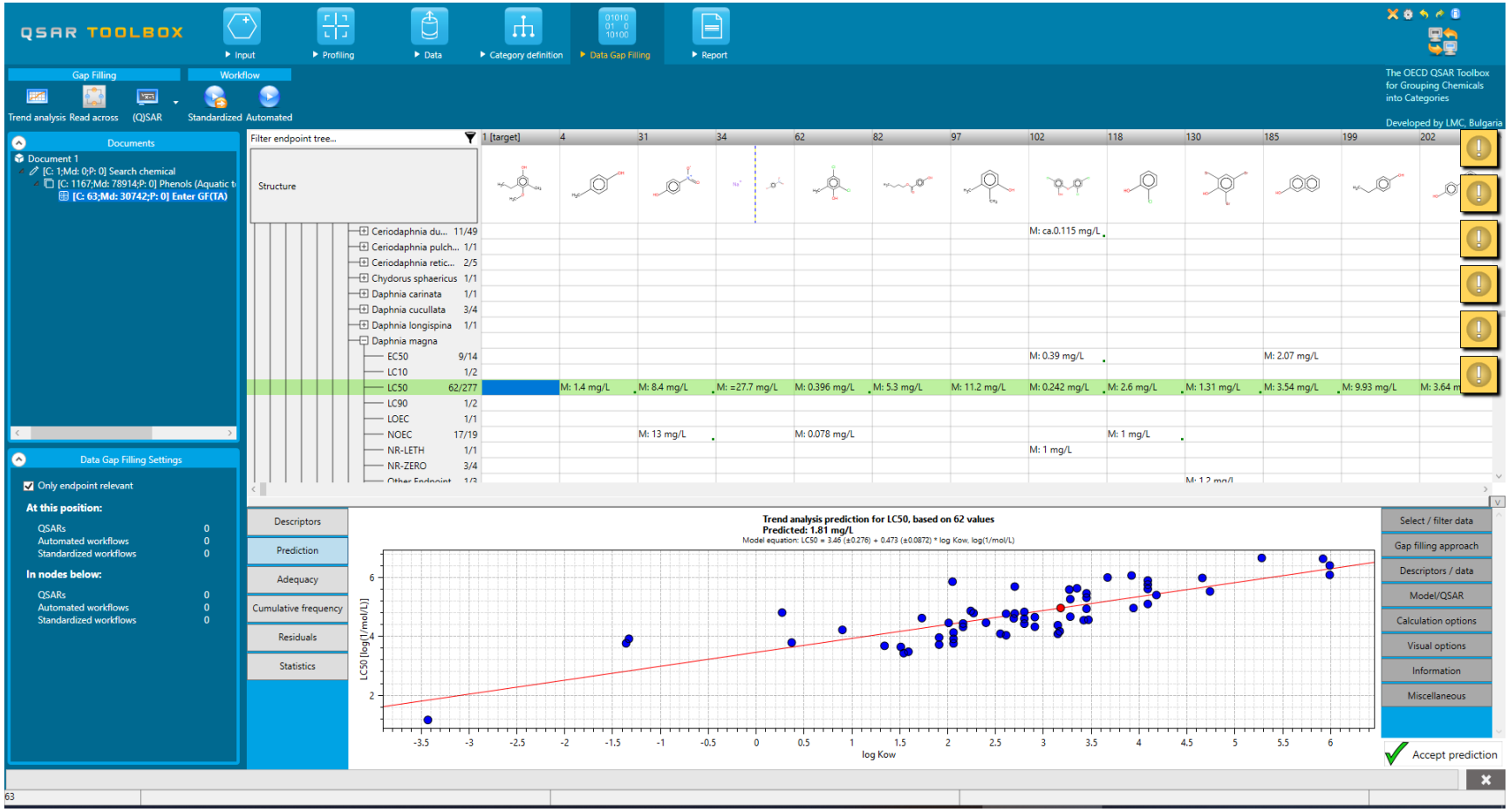
Data Gap Filling

Interpreting Trend analysis

- The resulting plot outlines the log of the experimental LC50 results of all analogues (Y axis) according to a descriptor (X axis) with Log *Kow* being the default descriptor (see next slide).
- The **RED** points represent the predicted value for the target chemical.
- The **BLUE** points represent the experimental results available for the analogues used in the trend analysis.
- The square-shaped signs in the right side of the data gap filling window are the so-called "helpers". The helpers are notifying messages that provide different type of information related to the used data points in the prediction.
- Before accepting the estimated result for the target chemical, the trend analysis should be further refined by subcategorization (see the next slides).

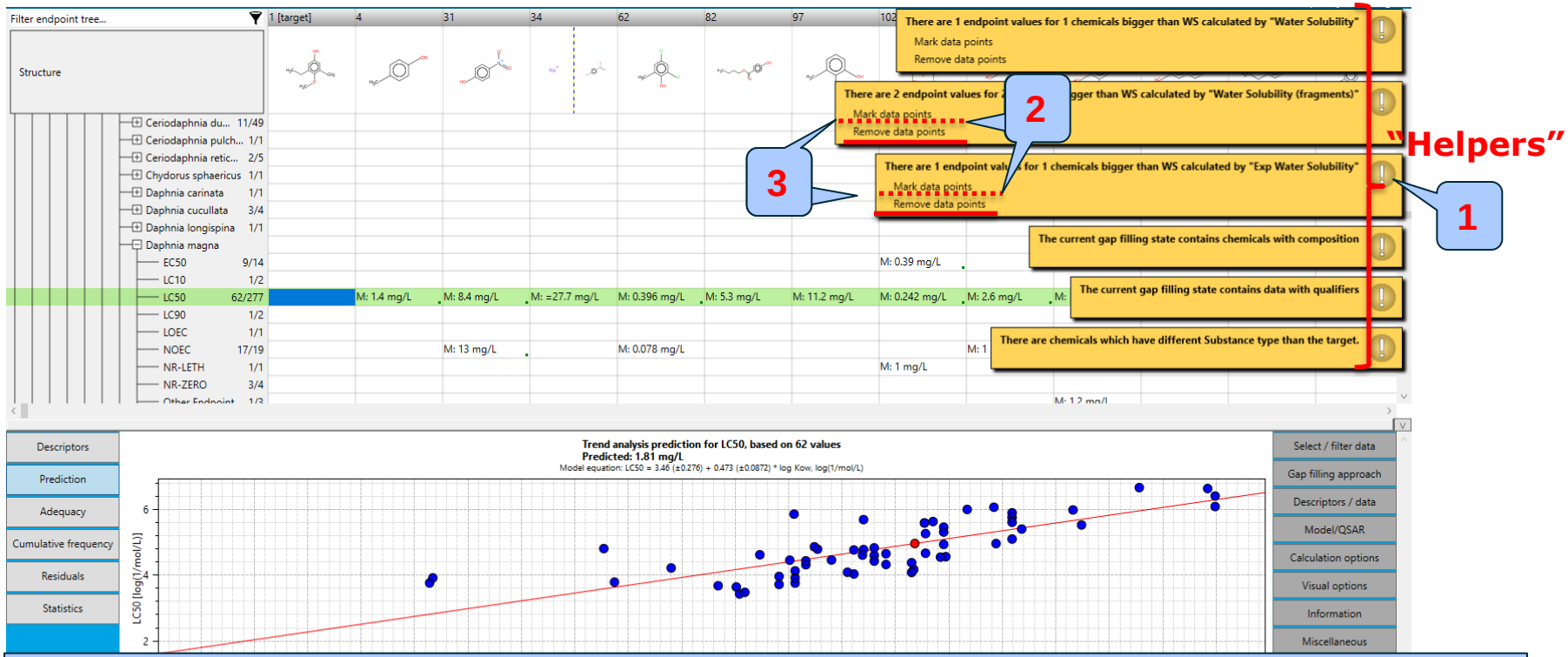
Data Gap Filling

Results of Trend analysis



Data Gap Filling

Results of Trend analysis



1. Click the helpers to see the information that they provide;
2. Click on "**Mark data points**" to mark the chemicals meeting the criteria explained in the helper (e.g. mark the chemicals which have 2 endpoint values (LC50) bigger than the water solubility (WS) of the chemical);
3. Click ***Remove data points*** in order to eliminate the data points that have values bigger than WS of the chemical (click on all helpers). Once the data points are removed the helpers disappear.

Data Gap Filling

Subcategorization

- Remember in the Toolbox, a category refers to a group of chemicals which have the same profiling result according to one of the profilers listed in the module *Profiling*.

- Subcategorization refers to the process of applying additional profilers to the previously defined category. The subcategorization identifies chemicals which have differing profiling results and eventually eliminating these chemicals from the final category.

Data Gap Filling

Subcategorization

In this example, subcategorization allows for the elimination of analogues which are dissimilar to the target chemical with respect to:

- OASIS Mode of action (all except phenols and anilines)

  The categorization based on mode of action identifies analogues having the same mode of action as the target which is in the group of phenols and anilines.

- Chemical elements

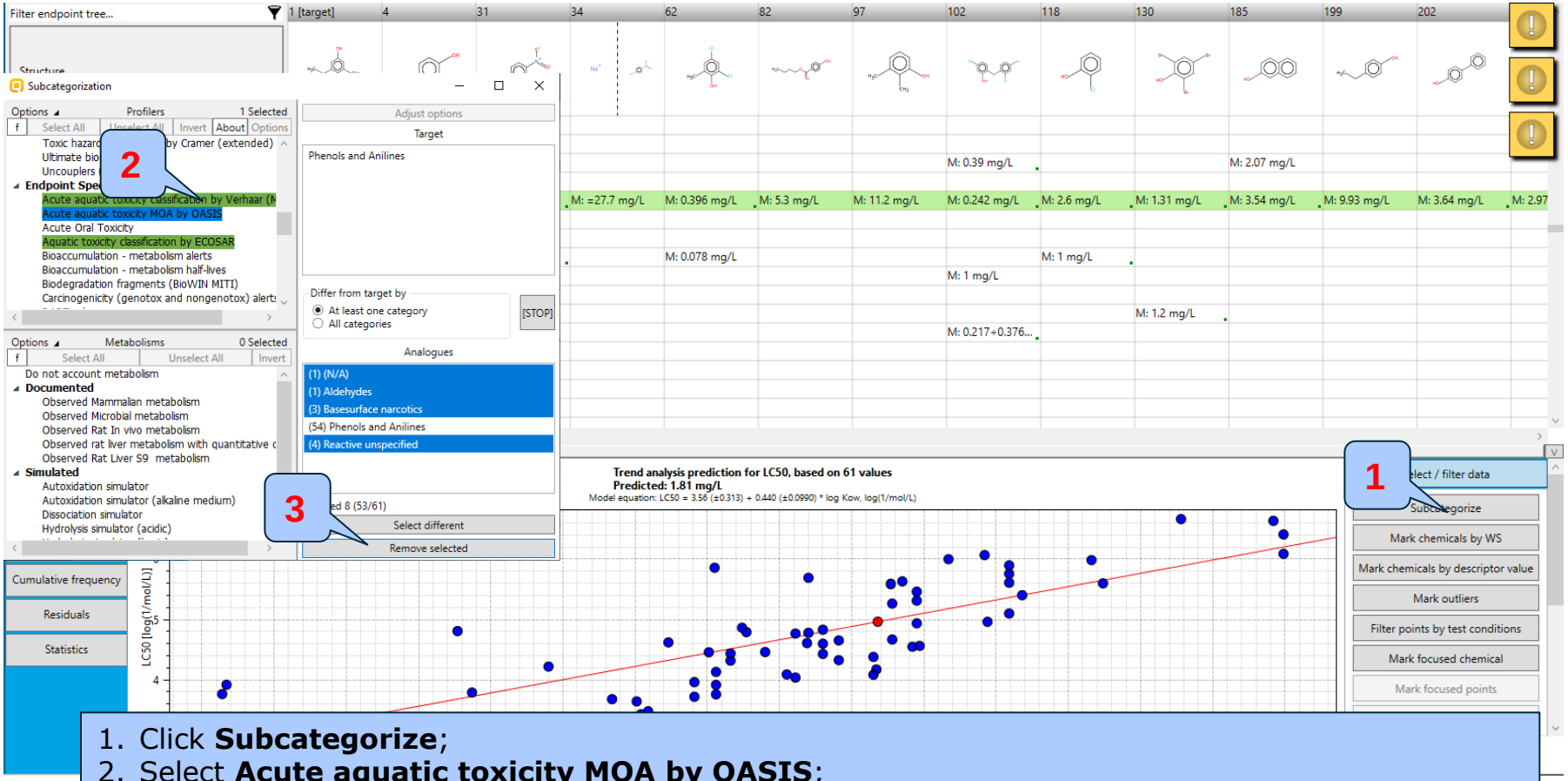
The profiler aimed to identify analogues consisting of same elements as those presented in the target chemical

Subcategorization is demonstrated in the next 2 slides.

Note: Expert judgement should always be used when removing chemicals.

Data Gap Filling

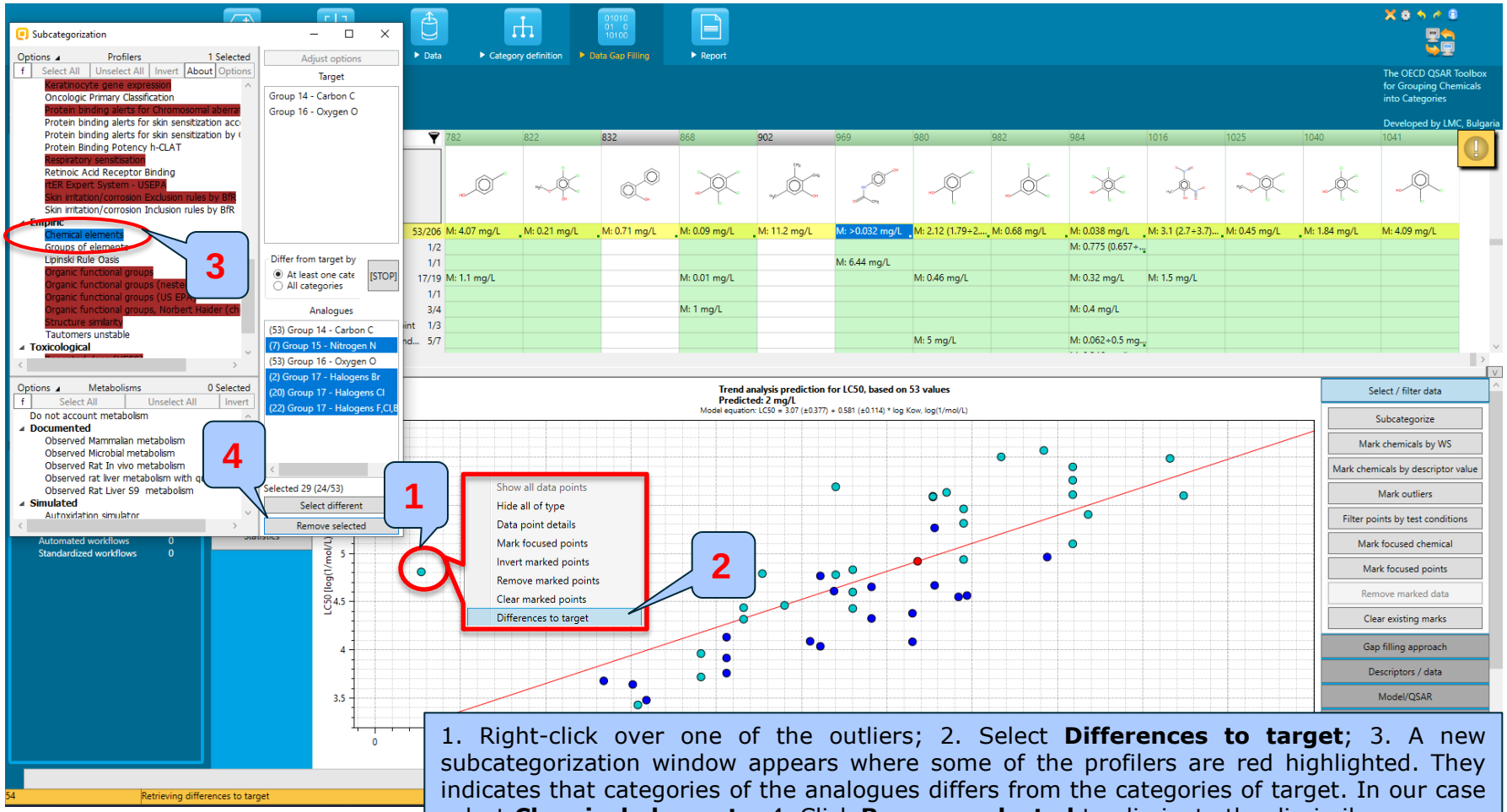
Subcategorization by Acute-aquatic toxicity MOA



1. Click **Subcategorize**;
2. Select **Acute aquatic toxicity MOA by OASIS**;
3. Click **Remove selected** to eliminate the dissimilar chemicals.

Data Gap Filling

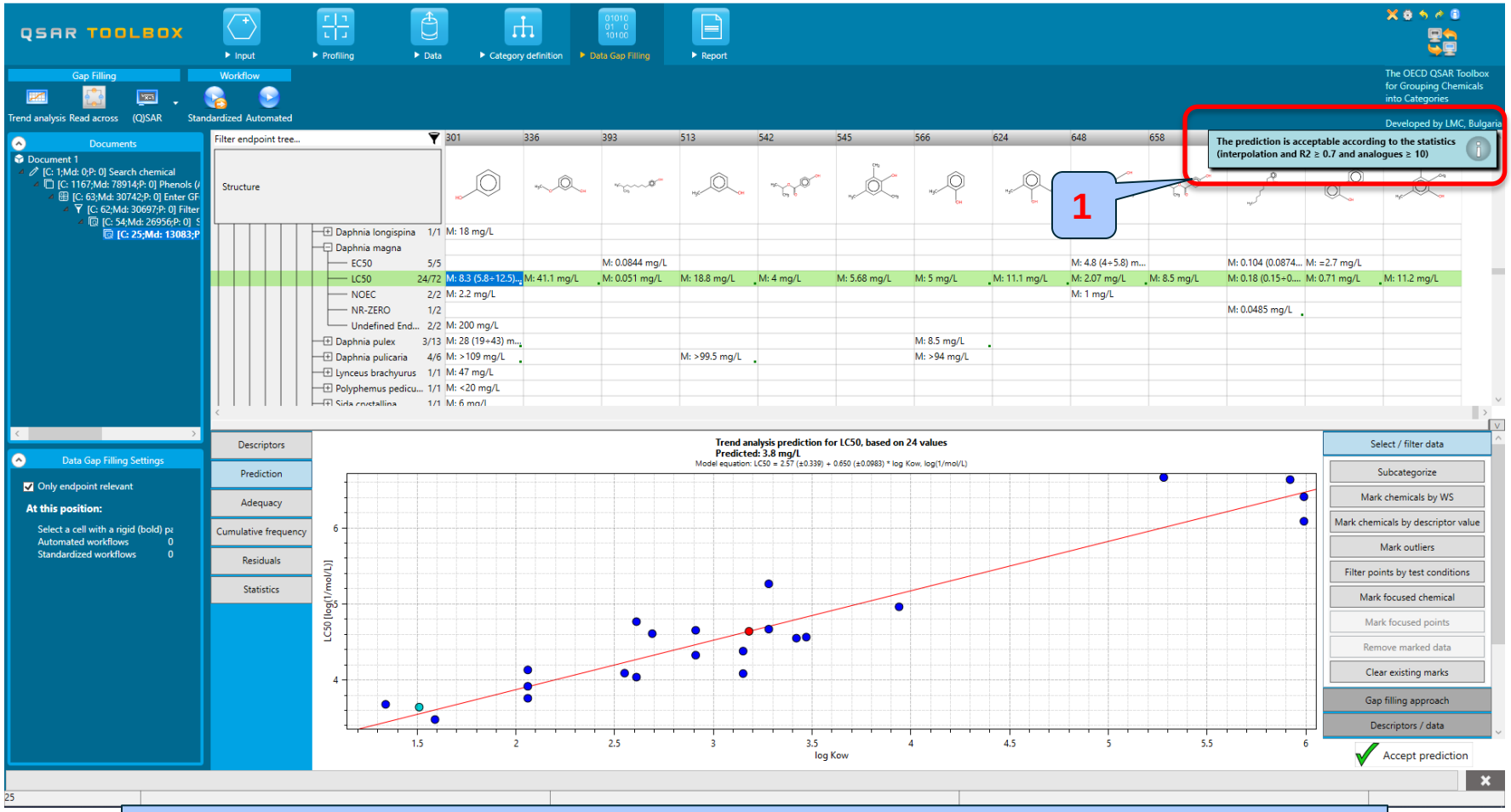
Subcategorization by Chemical elements



1. Right-click over one of the outliers; 2. Select **Differences to target**; 3. A new subcategorization window appears where some of the profilers are red highlighted. They indicates that categories of the analogues differs from the categories of target. In our case select **Chemical elements**; 4. Click **Remove selected** to eliminate the dissimilar

Data Gap Filling

Results



1. Helper indicating that the prediction is acceptable according to the statistic appears.

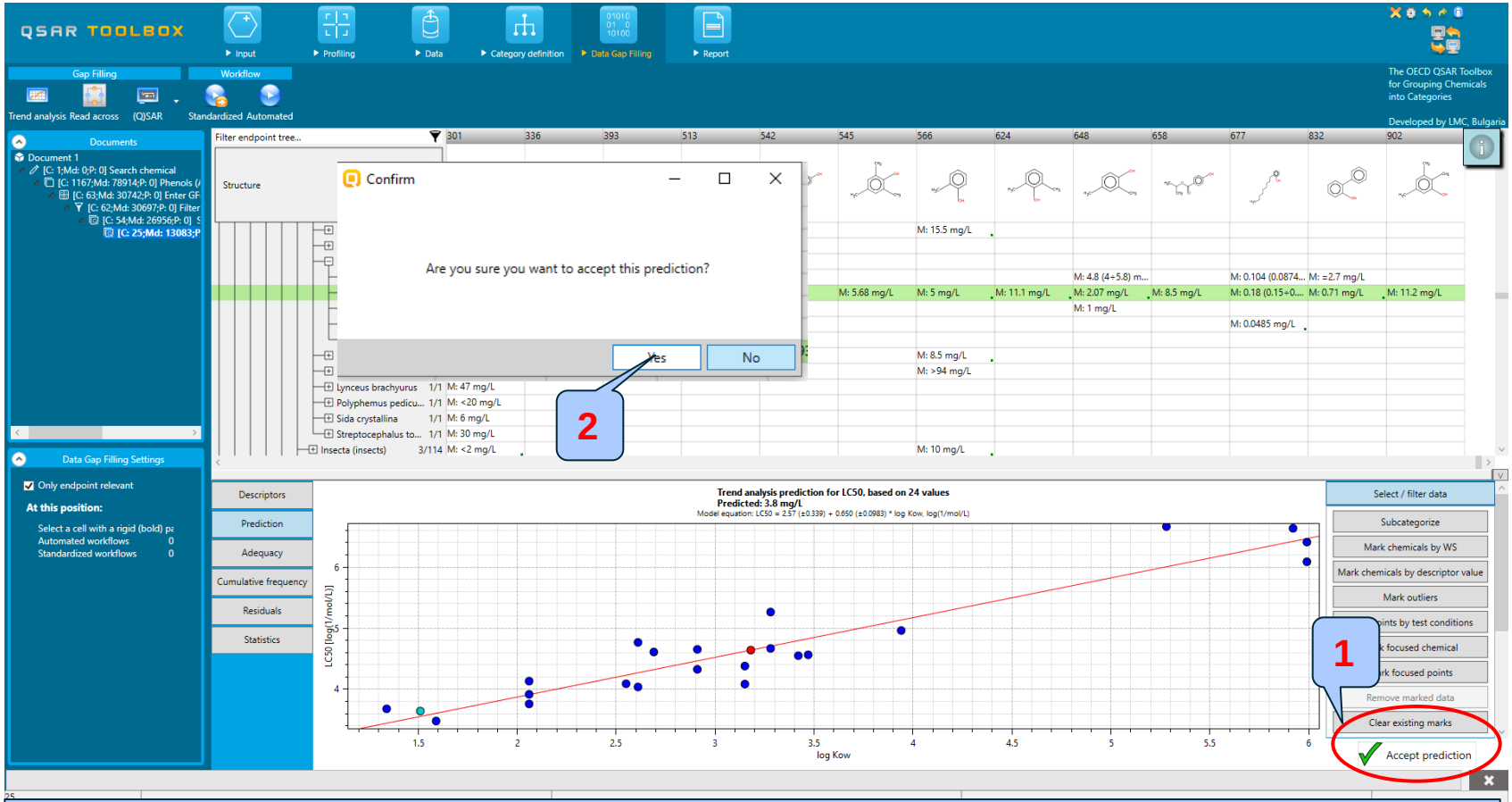
Data Gap Filling

Results

- All the chemicals remaining in the graph have a consistent profile relevant for aquatic toxicity (i.e. Substance type, Classification by ECOSAR, MOA by OASIS and Chemical elements) – they all are phenols
- By accepting the prediction the data gap is filled (see next screen shot).

Data Gap Filling

Accepting prediction result



1. Click **Accept prediction**;
2. Click **Yes** to confirm the prediction.

The prediction is accepted successfully and the system automatically returns you to the data matrix.

Outlook

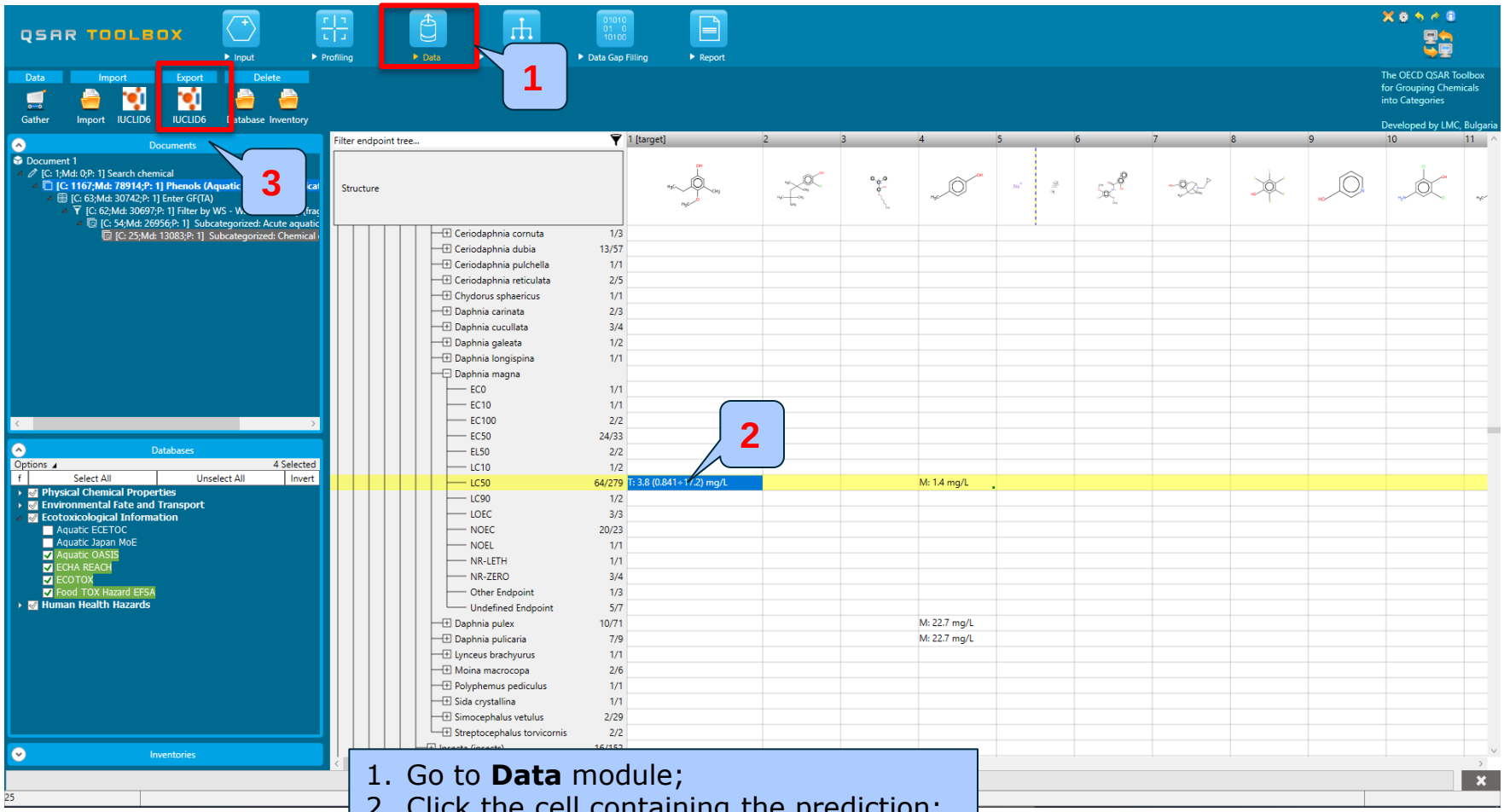
- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - Data
  - Category definition
  - Data Gap Filling
  - **Export a prediction to IUCLID6**

Export prediction to the IUCLID 6
Overview

- The OECD QSAR Toolbox allows the users to export predicted results (by means of the Filling Data Gap tools) to IUCLID 6.
- The way of exporting is connecting to an IUCLID 6 server (via WebServices) and assigning the predicted endpoint data to a selected substance.
- A wizard will guide the user through the different steps of exporting (see next screenshot).

Exporting the prediction to IUCLID 6

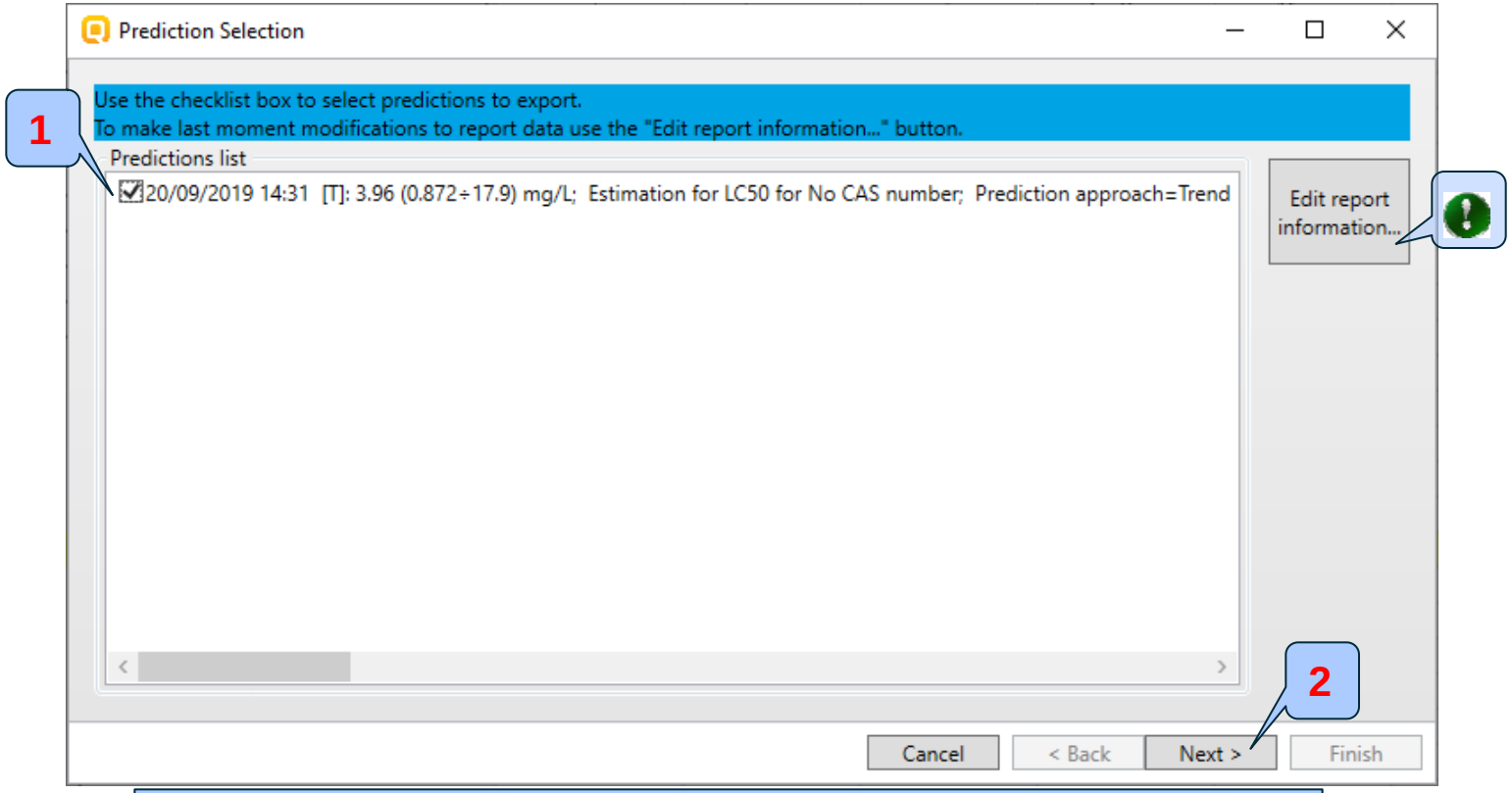
Case study



1. Go to **Data** module;
2. Click the cell containing the prediction;
3. Click **IUCLUD6**.

Exporting the prediction to IUCLID 6

Case study

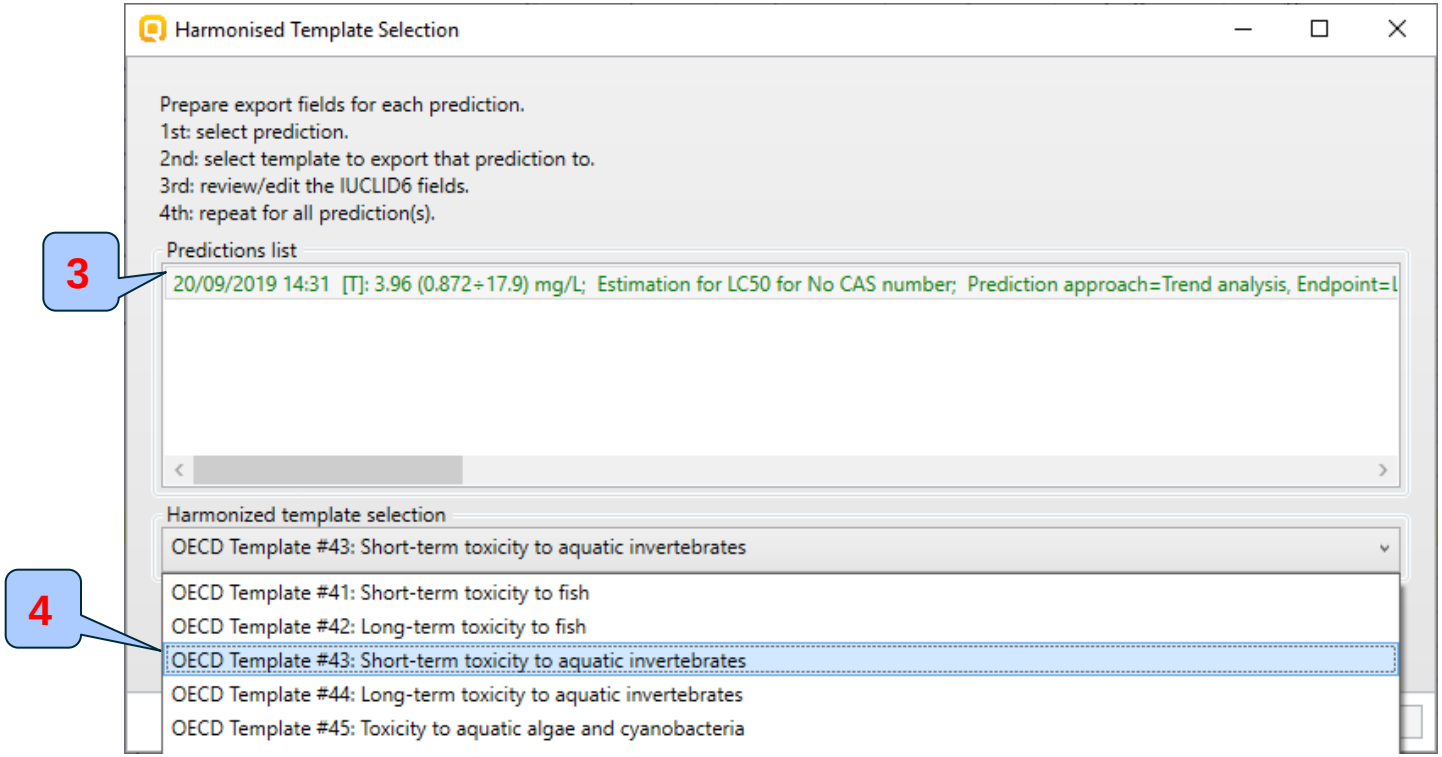


1. Tick the prediction to export;
2. Click **Next** to move through the next step of the export.

The user could also edit the report information.

Exporting the prediction to IUCLID 6

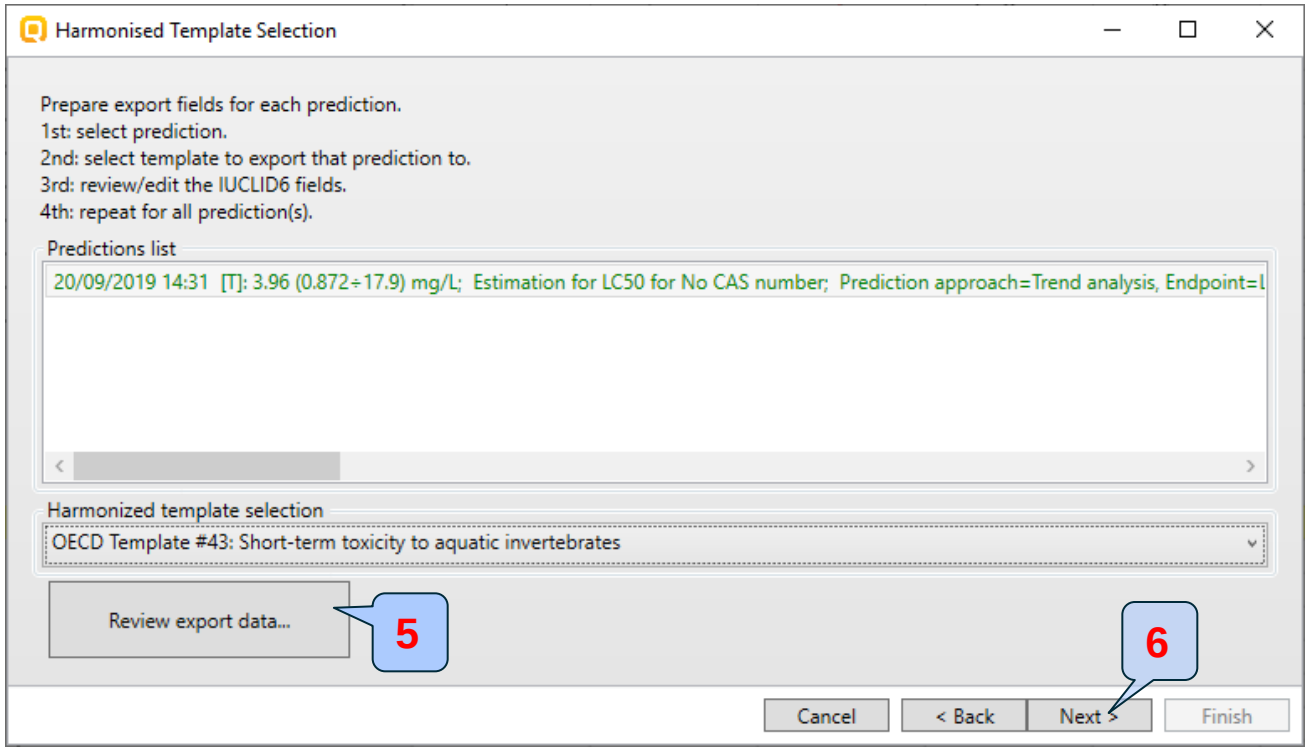
Case study



3. Select the prediction;
4. Select a template to export the prediction;

Exporting the prediction to IUCLID 6

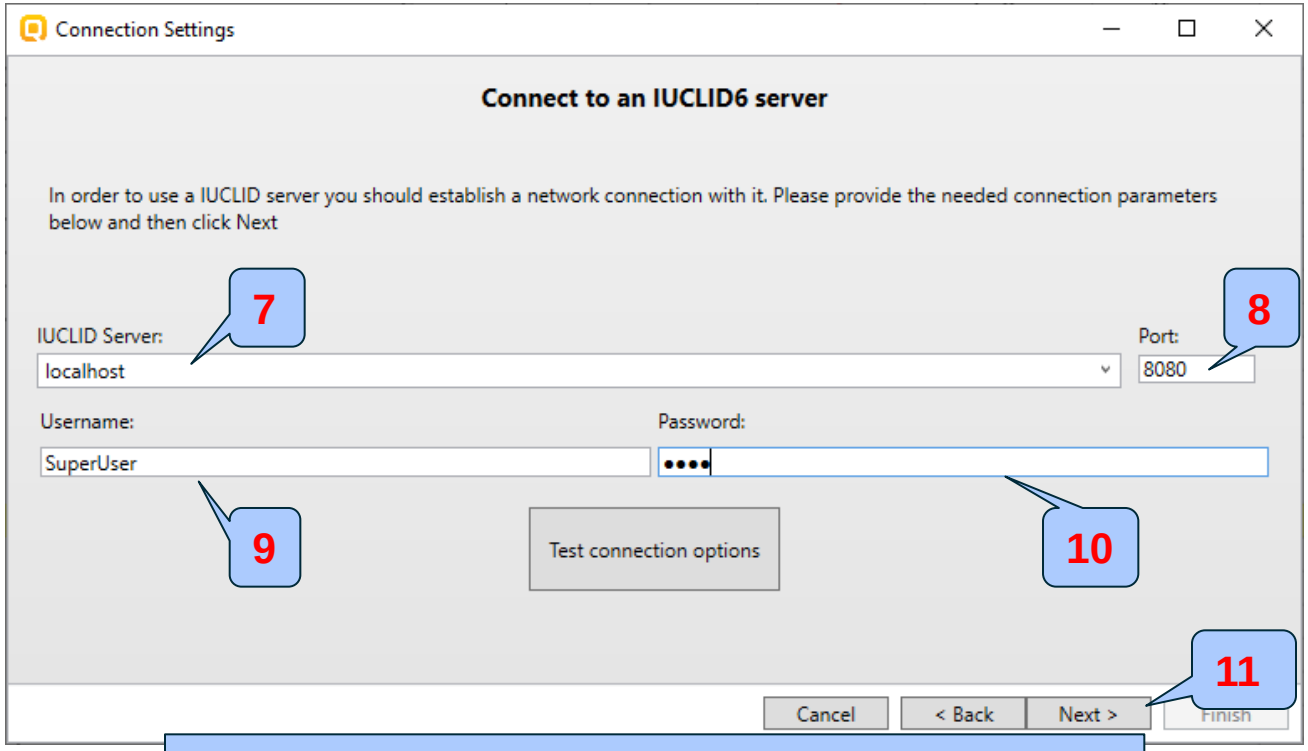
Case study



5. Review/edit the IUCLID6 fields;
6. Click **Next;**

Exporting the prediction to IUCLID 6

Case study



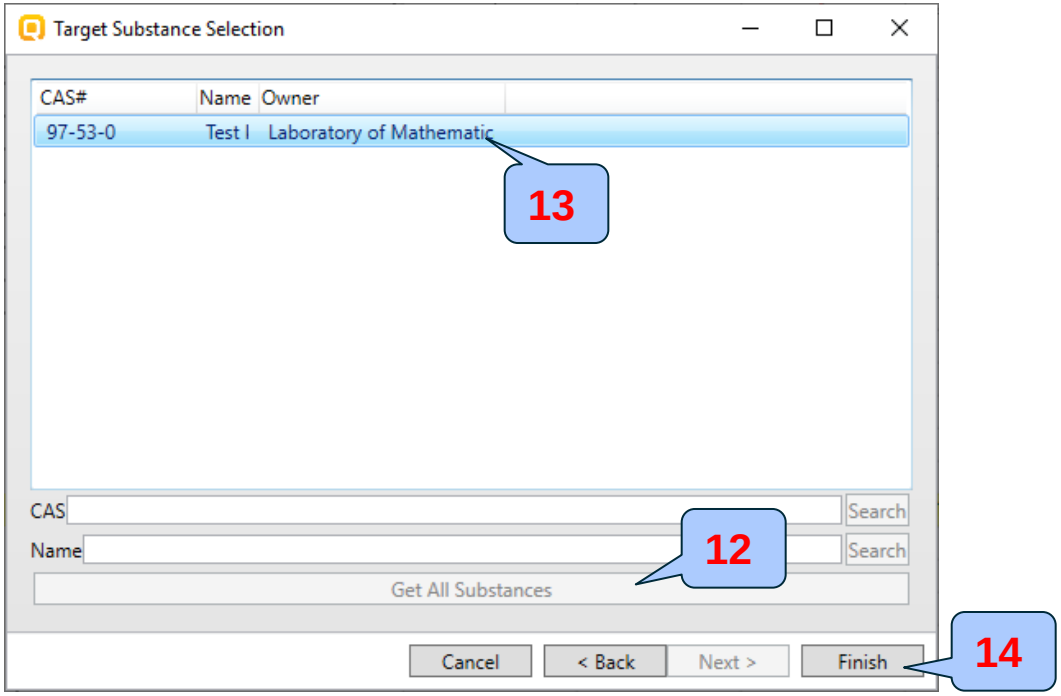
7. Write **IUCLID Server** name;
8. Fill in the **Port** number;
9. Fill in the **Username**;
10. Fill in the **password**: root;
11. Click **Next**.

Note:

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

Exporting the prediction to IUCLID 6

Case study



12. Click **Get All Substances**;
13. Select the chemical to which the prediction to be exported;
14. Click **Finish**.

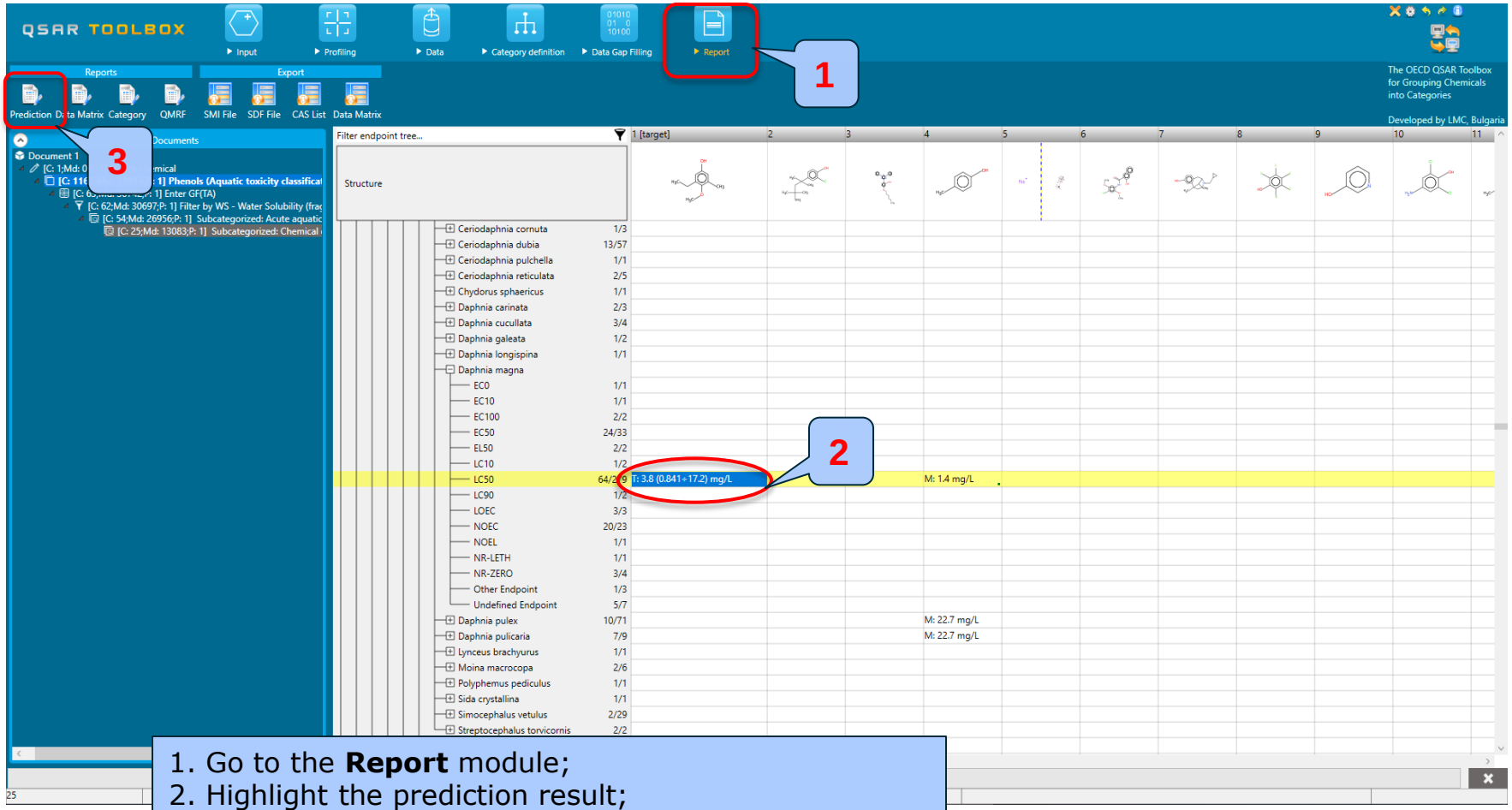
Outlook

- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow of the exercise**
  - Chemical Input
  - Profiling
  - Data
  - Category definition
  - Data Gap Filling
  - Export a prediction to IUCLID6
  - **Report**

Report
Overview

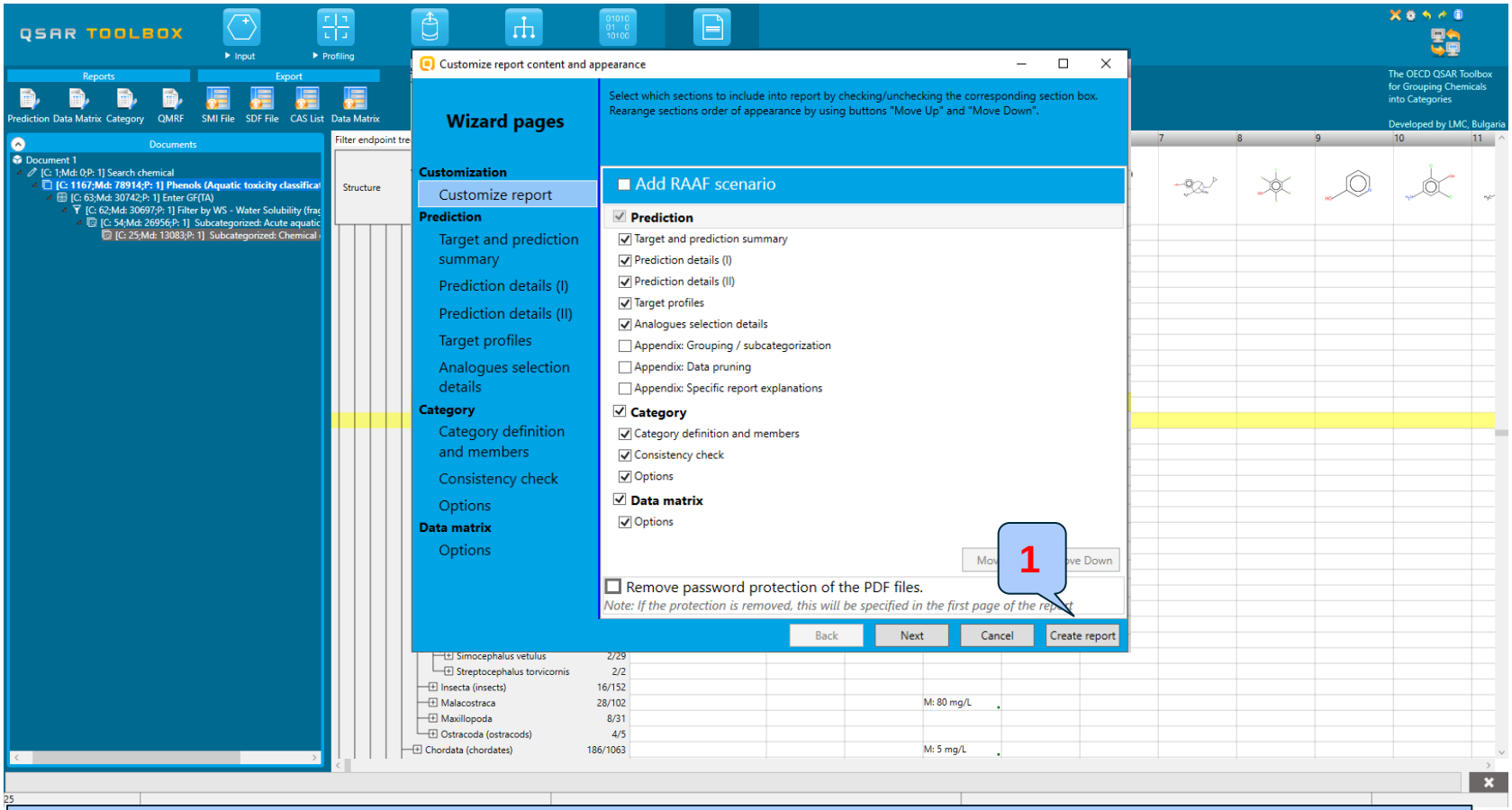
- The report module can generate a report on any of predictions performed with the Toolbox.
- The report module contains predefined report templates, which can be customized.
- The report can then be printed or saved in different formats.

Report
Generation report



1. Go to the **Report** module;
2. Highlight the prediction result;
3. Click **Prediction** to create a report.

Report
Generation report



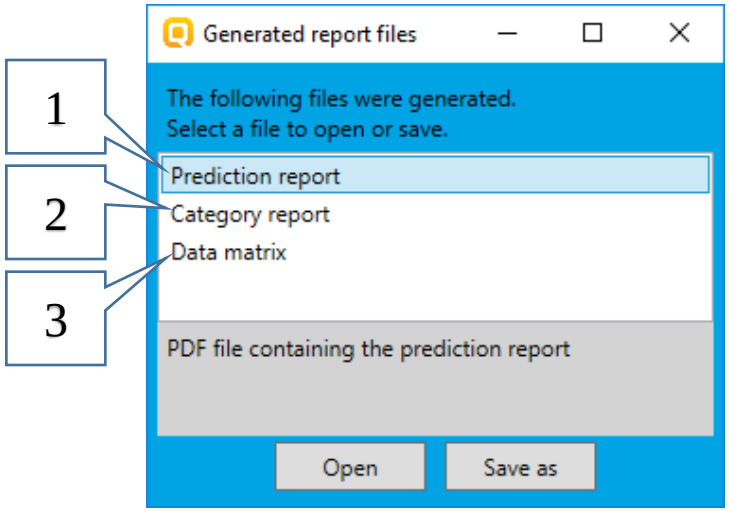
Report wizard pages appears, where the user could customize the report content and appearance. Some of the fields in the report are automatically populated by the system.

1. Click **Create report** to generate the report.

Report
Generation

After clicking *Create report* button, *Generated report files* window appears. It contains three type of files:

1) **Prediction report** - a PDF file containing the prediction information related to the target.
2) **Category report** - a PDF file containing information for the consistency of the final category (target plus used analogues)
3) **Data matrix** - a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.



Note: All generated files should be provided when submit a prediction.

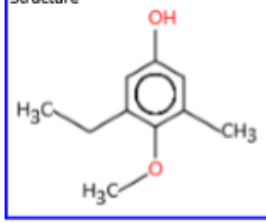
Report
Generated report files

Prediction report

QSAR Toolbox prediction for single chemical

Date: 3 Apr 2020
Author(s):
Contact details:

Target information

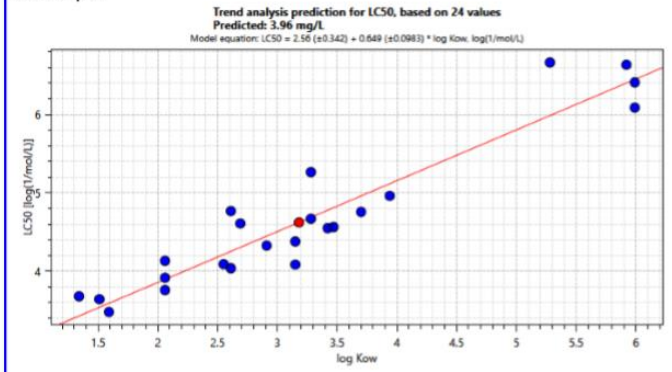
Structural information	Numerical identifiers	Chemical names
SMILES: CCc1cc(O)cc(C)c1OC Structure 	CAS#: No CAS number Other: N/A	

Prediction summary

Predicted endpoint: LC50; Mortality; Daphnia magna; 48 h; No guideline specified
Predicted value: 3.8 (from 0.841 to 17.2)
Unit/scale: mg/L
Data gap filling method: Trend analysis
Summary: *manually editable field*
Not provided by the user

Prediction details (I)

Predicted value: 4.62 log(1/mol/L), conf.range: (3.97 ; 5.28) at 95.0%
Predicted endpoint (OECD Principle 1 - Defined endpoint): Ecotoxicological Information -> Aquatic Toxicity -> Mortality -> LC50 -> 48 h -> Animalia (animals) -> Arthropoda (arthropods) -> Branchiopoda (branchiopods) -> Daphnia magna
Prediction plot:



Calculation approach (OECD principle 2 - Unambiguous algorithm): Linear approximation
Model equation: LC50 = 2.56 (±0.342) + 0.649 (±0.0983) * log Kow, log(1/mol/L)
Active descriptor: log Kow (calculated)
Data usage: Arithmetic mean (average) value*
Statistics of the prediction model:
N = 24; count of data points
R2 = 0.895; coefficient of detemination
R2adj = 0.890; adjusted coefficient of detemination
SSR = 2.12; sum of squared residuals
s = 0.297; sample standard deviation of residuals
F = 187; Fisher function
*When multiple values are available for the same chemical, their arithmetic mean (average) value is taken in prediction calculations

Report
Generated report files

Category report

QSAR Toolbox report for category

1. Category definition

1.1. Category definition

Category name *manually editable field*

Not provided by the user

Covered (target) endpoint(s)

- Ecotoxicological Information/Aquatic Toxicity: LC50, Daphnia magna, Branchiopoda (branchiopods), Arthropoda (arthropods), Animalia (animals), Mortality, Duration=48 h

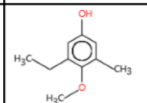
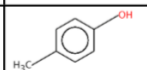
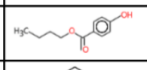
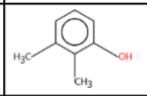
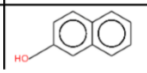
Category hypothesis *manually editable field*

Not provided by the user

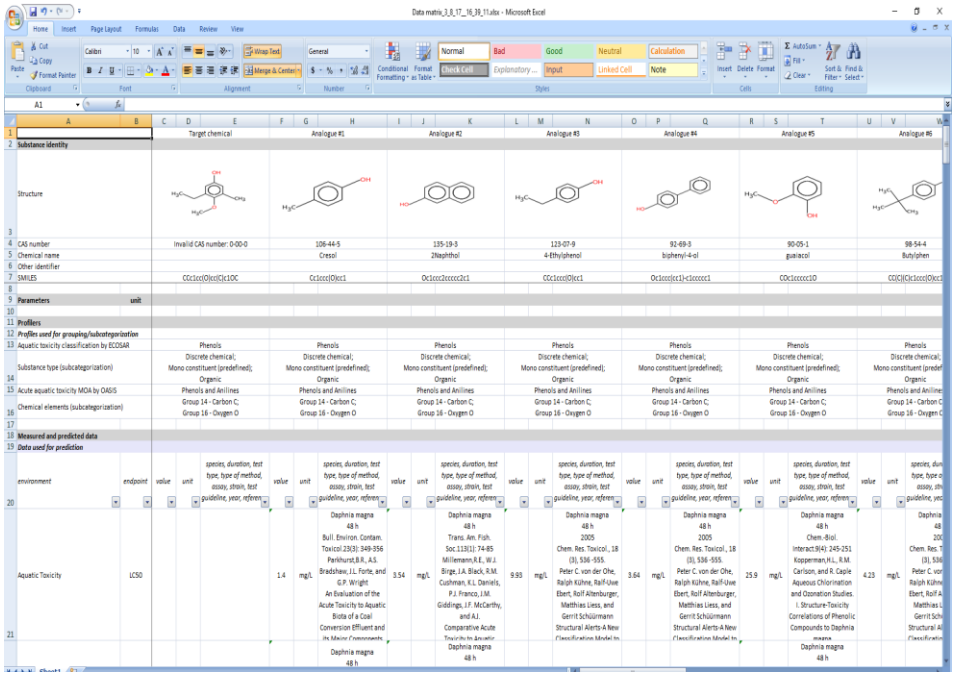
1.2. Category members

Information of category members

Table of category members

#	CAS	Name	SMILES	Structure
1	No CAS number		CCc1cc(O)cc(C)c1OC	
2	106-44-5	Cresol	Cc1ccc(O)cc1	
3	94-26-8	Butylparaben	CCCCOC(=O)c1ccc(O)cc1	
4	526-75-0	2,3-xylenol	Cc1cccc(O)c1C	
5	135-19-3	135-19-3	Oc1ccc2ccccc2c1	

Data matrix report



Outlook

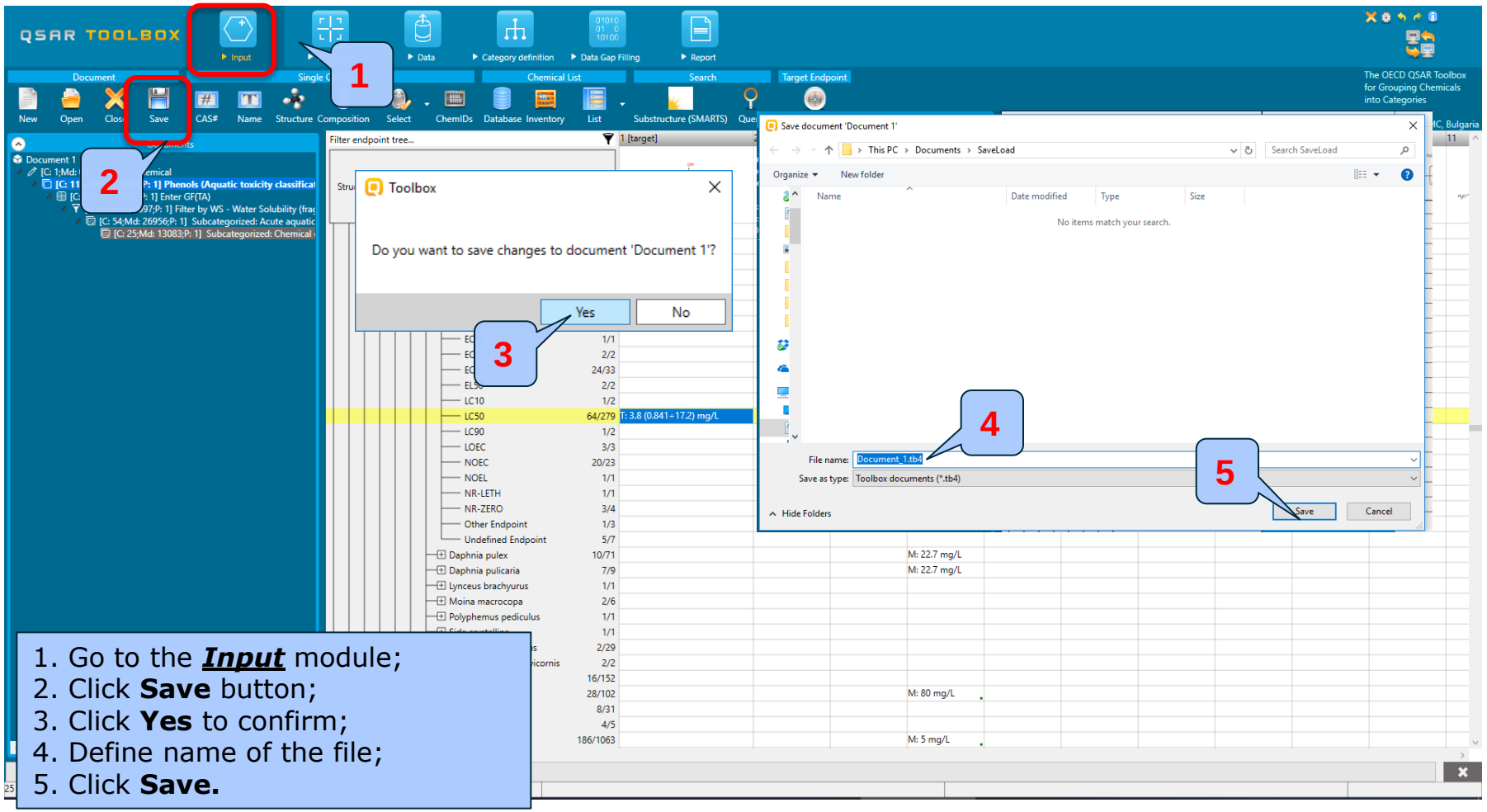
- Background
- Keywords
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow of the exercise
- **Save the prediction result**

Saving the prediction result

- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc., on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).

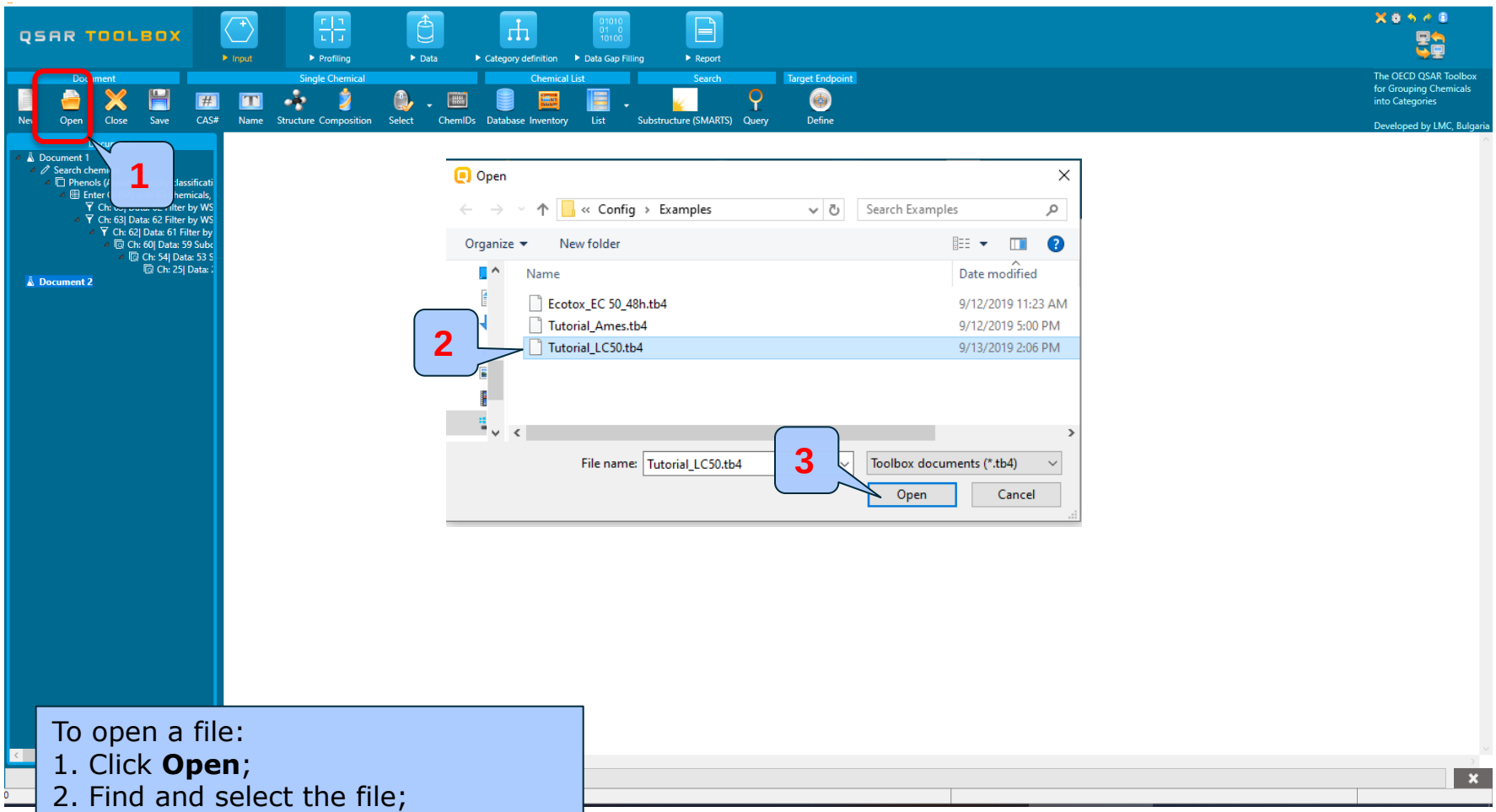
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result



1. Go to the ***Input*** module;
2. Click **Save** button;
3. Click **Yes** to confirm;
4. Define name of the file;
5. Click **Save.**

Open file



To open a file:
1. Click **Open**;
2. Find and select the file;
3. Click **Open**.

Congratulations!

- You have now been introduced to the work flow of the Toolbox and completed the tutorial on data gap filling by trend analysis and exported the prediction to IUCLID 6.
- You have been introduced to the six modules of the Toolbox, the basic functionalities within each module and the rationale behind each module.
- Remember, proficiency comes with practice!