

OECD QSAR Toolbox v.3.4

Step-by-step example of how to categorize an inventory by mechanistic behaviour of the chemicals which it consists

Outlook

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

Background

- This is a step-by-step presentation designed to take Toolbox users through the workflow of categorizing an inventory using mechanistic similarity.

Outlook

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

Objectives

This presentation demonstrates a number of functionalities of the Toolbox including:

- Clustering chemicals from an imported inventory (chemical list).
- Selecting a cluster and place it into separate data matrix for later analyses.
- And, apply a trend analysis for aquatic toxicity to the selected cluster.

Outlook

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

Specific Aims

- To examine the workflow of clustering an inventory.
- To introduce the user to new functionalities within selected modules.
- To explain the rationale behind each step of the exercise.

Outlook

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

Trend analysis

- For a given 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 structural or chemical property.
- A demonstration of consistent trends in the behavior of a group of chemicals is one of the desirable attributes of a chemical category and an indicator that there is a common mechanism of action. When a consistent trend is observed and some chemicals in the category have measured values, missing values can be estimated by simple scaling from the measured values.

Outlook

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

The Exercise

Part I

- In this part of the exercise we will import a chemical list containing 300 substances, representing a typical inventory of chemicals used by a down stream user.
- The inventory will be loaded into the Toolbox data matrix.
- The inventory will be profiled by the “Aquatic toxicity classification by ECOSAR Classification” profiling scheme.
- This inventory will be clustered by the ECOSAR Classification grouping method.

The Exercise

Part II

- In this part of the exercise an example cluster or defined category will be examined.
- Data on the chemicals which belong to selected cluster will be gathered.
- A data gap will be filled by trend analysis - In this exercise we will predict the toxicity towards mortality of the *Pimephales promelas* (LC50 for 24h).
- A report of the prediction will be generated.

The Exercise

Necessary to download the file

- The inventory is stored in the following smi file: OECD Mock Inventory.smi*
- This file contains a list of Chemical Abstract Service (CAS) numbers, which can be opened with Notepad.

*OECD Mock Inventory.smi - file is packed in the zip file with all power point tutorials posted on LMC website

Also the file could be found in the example folder of Toolbox installation

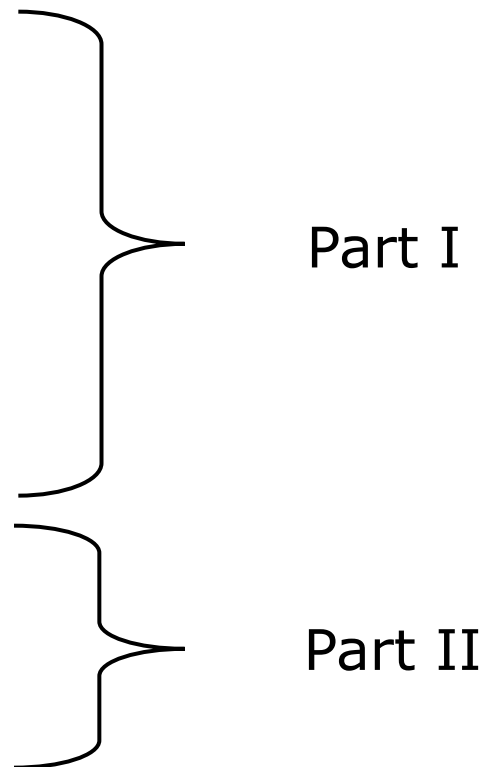
Outlook

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

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**

- Chemical Input
- Profiling
- Endpoints
- Category Definition
- Filling Data Gaps
- Report



Outlook

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

Chemical Input Overview

- This module provides the user with 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 is the correct one.

Chemical Input

Ways of Entering a Chemical

User Alternatives for Chemical ID:

A. Single target chemical

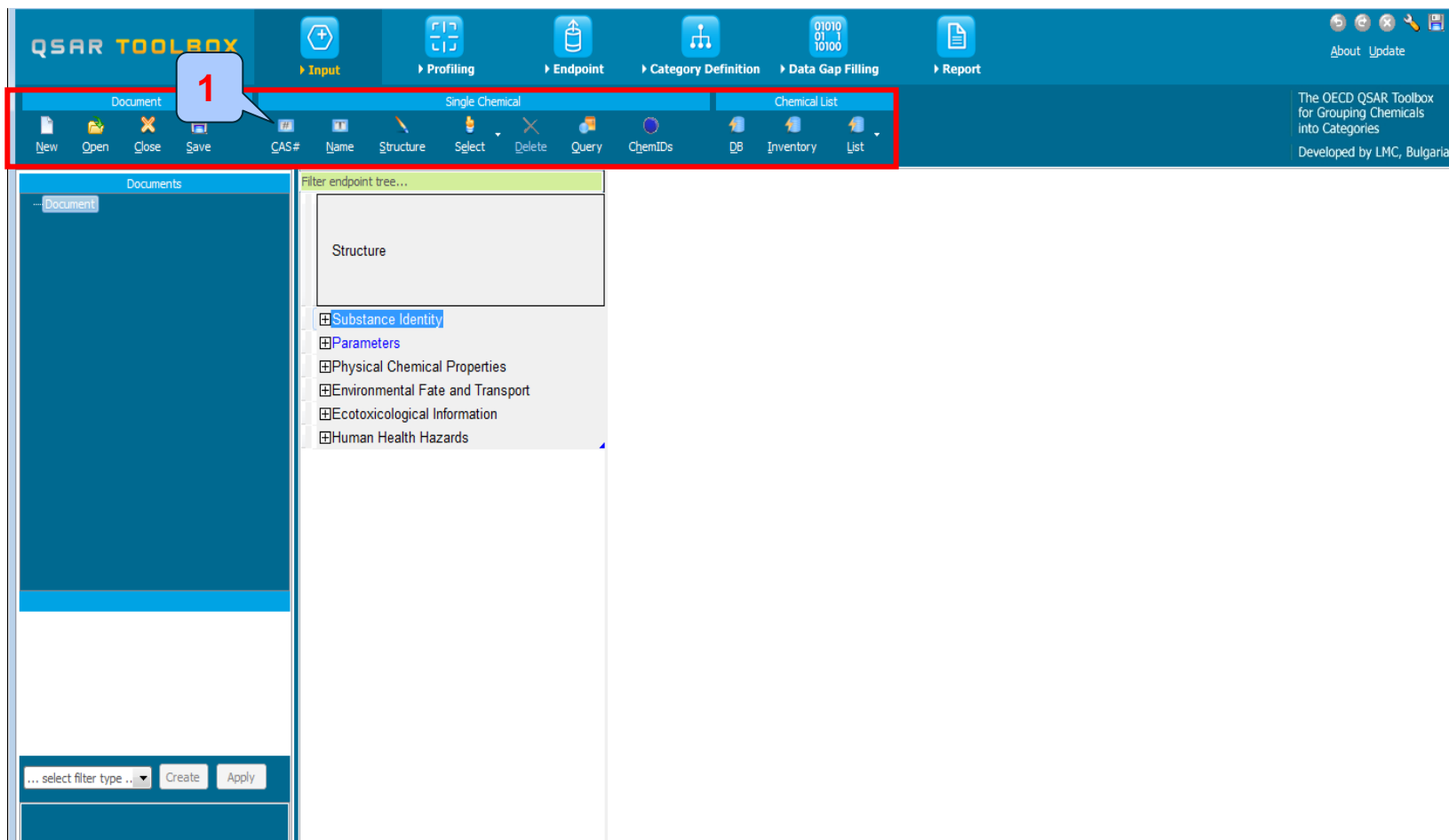
- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Chemical IDs such as EC number, Einescs number

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Chemical Input Screen

Input target chemical by CAS#



Chemical Input Screen

Input target chemical by CAS#

The screenshot shows the QSAR Toolbox interface with the 'Chemical List' button highlighted by a red box and callout '1'. An 'Open' dialog box is open, showing a file list with 'OECD Mock Inventory.smi' selected, indicated by callout '2'. The 'Open' button in the dialog is highlighted by callout '3'. A callout '4' points to a folder named 'From examples folder' in the 'Chemical List' dropdown menu.

1. Click on List **2. Select** OECD Mock Inventory.smi **3. Click** Open
4. The file is placed also in the example folder. The example folder could be opened from list button

Chemical Input Screen

Missing SMILES

- Since in the entered list there are CAS numbers for which there are no SMILES in the Toolbox, the user will be asked if he/she wants the Toolbox to search for in the databases. In this example select **NO**.

Chemical Input Screen

Missing SMILES

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below the menu is a toolbar with icons for 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Select', 'Delete', 'Query', 'ChemIDs', 'DB', 'Inventory', and 'List'. A 'Documents' panel on the left shows a 'Document' tab. The main workspace is titled 'Filter endpoint tree...' and contains a tree view with categories such as 'Structure', 'Substance Identity', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazards'. A 'Confirm' dialog box is overlaid on the workspace, containing the text: 'There are empty SMILES with defined CAS in the list. Would you like to search in the database with CAS?'. The dialog has 'Yes' and 'No' buttons. A red '1' in a blue box points to the 'No' button. A text box at the bottom right of the screenshot contains the instruction: '1. Click NO in dropdown box.'

Chemical Input Screen

Displaying structures

- The structures of the 300 chemicals are displayed with 2D image. The missing SMILES are displayed with empty window - for example chemical #103 (see next screen shot).
- These empty windows remind the user where are the gaps in the inventory.
- Remember all subsequent functions are based on chemical structure, so the chemicals with blank windows will not be placed in a cluster or category.

Chemical Input Screen

Displaying structures

The screenshot displays the QSAR Toolbox interface. At the top, there is a menu bar with options: Document, Single Chemical, and Chemical List. Below this is a toolbar with icons for New, Open, Close, Save, CAS#, Name, Structure, Select, Delete, Query, ChemIDs, DB, Inventory, and List. The main workspace is a grid of 10 columns, each representing a target (97 to 105). Each column contains a chemical structure. The structure for target 103 is highlighted with a red box. On the left side, there is a sidebar with a 'Documents' list and a 'Filter endpoint tree...' section. The 'Structure' tab is selected in the filter tree, and a list of endpoints is shown: Substance Identity, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. At the bottom left, there is a search bar with the text 'Cc1cc(Br)cc(C)c1O' and a corresponding chemical structure. The status bar at the bottom shows '300 Document' and '1/0/0'.

Chemical Input

Chemical Identity

- **Double click** "Substance Identity"; this displays the chemical identification information.
- Note that existing in the Toolbox names of target chemical are in different colors.
- The workflow on the first module is now complete, and the user can proceed to the next module.

Chemical Input Chemical Identity

The screenshot displays the QSAR Toolbox interface with the 'Chemical Identity' section active. A table lists chemical entries across seven columns, with a red circle highlighting the 'Chemical Name' and 'Molecular Formula' rows for entries 1 through 6.

Filter endpoint tree...	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]
Structure							
Substance Identity	2374-05-2	2426-08-6	3680-02-2	4837-01-8	5390-04-5	100-00-0	100-1
CAS Number	EINECS:2191531	EINECS:2193764	EINECS:2229580	EINECS:2254224	EINECS:2263836	EINECS:2026006	EINECS:2026006
Chemical IDs	4-bromo-2,6-dimethylphenol	1-butoxy-2,3-epoxypropane, 2-(butoxypropyl)glycidyl ether	methyl vinyl sulfone (methylsulphonyl)ethane	[(2-bromoethyl)thio]ethane	4-pentyn-1-ol	1-chloro-4-nitrobenzene	4-nitrophenol
Chemical Name	4-bromo-2,6-dimethylphenol	butyl glycidyl ether				1-chloro-4-nitrobenzene	4-nitrophenol
Molecular Formula	C8H9BrO	C11H20O2	C3H6O2S	C8H9BrS	C5H8O	C6H4ClNO2	C7H7NO2
Structural Formula	Cc1cc(Br)cc(C)c1O	CCCCOCC1CO1	CS(=O)C=C	BrCCSc1ccc1	C#CCCO	O=N(O)c1ccc(Cl)cc1	O=N(O)c1ccc(O)cc1
Parameters							
Physical Chemical Properties							
Environmental Fate and Transport							
Ecotoxicological Information							
Human Health Hazards							

Chemical Input

Chemical Identity

The color code indicates the reliability of the chemical identifier:

- **Green:** There is a high reliability between the identifier and the structure. This color is attributed if the identifier is the same in several quality assured databases.
- **Yellow:** There is only a moderate reliability between the identifier and the structure. The color is attributed if the identifier is the same in several databases for which the quality assurance could not be established.
- **Red:** There is a poor reliability between the identifier and the structure. The color is attributed if the identifier is allocated to different structures in different databases.
- **Black:** The quality of the structure and its identifier is unknown.

Outlook

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

Profiling

Overview

- As you are aware the “profiling” refers to the electronic process of retrieving relevant information on the target compound(s), which are stored in the Toolbox other than fate and toxicity data.
- Key to this exercise is the information on likely mechanism(s) of action.

Profiling

Side-Bar of Profilers

- Remember detailed explanation of the different profilers is available in the in Manual for getting started (Chapter 4). <http://www.oecd.org/dataoecd/58/56/46210452.pdf>
- Background information can be retrieved by highlighting a profiler and a click on the button "View".
- For grouping chemicals into meaningful categories for aquatic toxicity the "Aquatic toxicity classification by ECOSAR" is one of the most adequate scheme.

Profiling Side-Bar of Profilers

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' side-bar is open, showing a list of profiling methods. A red circle with the number '1' highlights the 'Aquatic toxicity classification by ECOSAR' method. A red circle with the number '2' highlights the 'View' button in the side-bar's toolbar. Below the side-bar, a table shows the results of profiling for eight different chemical structures. The table columns are labeled '1 [target]' through '8 [target]'. The first column is expanded to show the 'ECOSAR Class Definition: Acid Halides' window, which includes a list of chemical structures and their SMILES string identifications.

Structure	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]	8 [target]												
Substance Identity																				
— CAS Number	2374-05-2	2426-08-6	3680-02-2	4837-01-8	5390-04-5	100-00-5	100-11-8	100-12-9												
— Chemical IDs	EINECS:2191531 4-bromo-2,6-dimet	EINECS:2193764 1-butoxy-2,3-epoxy	EINECS:2229580 methylvinyl sulfone	EINECS:2254224 1-(2-bromoethyl)thio	EINECS:2263836 4-pentun-1-ol	EINECS:2028096 1-chloro-4-nitroben...	EINECS:2028206 4-nitrobenzyl bromide	EINECS:2028206 4-ethylnitrobenzene												
ECOSAR Class Definition: Acid Halides	<p>The Acid Halide class is identified by any of the following structures:</p> $\begin{matrix} \text{O} \\ \parallel \\ \text{C}-\text{Halogen} \end{matrix} \quad \begin{matrix} \text{O} \\ \parallel \\ \text{P}-\text{Halogen} \end{matrix} \quad \begin{matrix} \text{O} \\ \parallel \\ \text{S}-\text{Halogen} \end{matrix}$ <p>The Halogens include chlorine (Cl), bromine (Br), fluorine (F) and iodine (I).</p> <p>The C(=O) carbon, P(=O) phosphorus and S(=O) sulfur can have other attachments. In the current ECOSAR program, these other attachments can be anything.</p> <p>SMILES String Identifications:</p> <table border="0"> <tr> <td>C(=O)Br</td> <td>C(=O)Cl</td> <td>C(=O)F</td> <td>C(=O)I</td> </tr> <tr> <td>P(=O)Br</td> <td>P(=O)Cl</td> <td>P(=O)F</td> <td>P(=O)I</td> </tr> <tr> <td>S(=O)Br</td> <td>S(=O)Cl</td> <td>S(=O)F</td> <td>S(=O)I</td> </tr> </table>								C(=O)Br	C(=O)Cl	C(=O)F	C(=O)I	P(=O)Br	P(=O)Cl	P(=O)F	P(=O)I	S(=O)Br	S(=O)Cl	S(=O)F	S(=O)I
C(=O)Br	C(=O)Cl	C(=O)F	C(=O)I																	
P(=O)Br	P(=O)Cl	P(=O)F	P(=O)I																	
S(=O)Br	S(=O)Cl	S(=O)F	S(=O)I																	
	C6H4ClNO2	C7H6BrNO2	C8H9NO2																	
	O=N(=O)c1ccc(Cl)cc1	O=N(=O)c1ccc(Br)cc1	CCc1ccc(N=O)cc1																	

1. Highlight the "Aquatic toxicity classification by ECOSAR" profiler; 2. Click View to see the drop-down box.

Profiling Side-Bar of Profilers

The screenshot displays the QSAR Toolbox software interface. The top toolbar includes buttons for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' button is highlighted with a callout box containing the number '2'. Below the toolbar, the 'Profiling methods' sidebar is visible, with a red circle around the 'Acute aquatic toxicity classification by ECOSAR' option. A callout box with the number '1' points to a green tick mark next to this option. The main window shows a table of chemical profiles with columns for CAS Number, Chemical IDs, Chemical Name, Molecular Formula, and Structural Formula. The table contains 8 columns of data, each representing a different chemical profile. The 'Acute aquatic toxicity classification by ECOSAR' option is checked, and the 'Apply' button is highlighted.

Filter endpoint tree...	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]	8 [target]	
Structure									
Substance Identity	2374-05-2	2426-08-6	3680-02-2	4837-01-8	5390-04-5	100-00-5	100-11-8	100-12-9	
CAS Number	EINECS:2191531	EINECS:2193764	EINECS:2229580	EINECS:2254224	EINECS:2263836	EINECS:2028096	EINECS:2028206	EINECS:2028206	
Chemical IDs	4-bromo-2,6-dimet... 4-bromo-2,6-dimet... 4-bromo-2,6-xylenol phenol, 4-bromo-2,...	1-butoxy-2,3-epox... oxirane, 2-(butoxy... bge butyl glycidyl ether glycidyl n-butyl ether n-butyl glycidyl ether oxirane, (butoxym... butyl 2,3-epoxypro... 2-(butoxymethyl)ox... butyl glycidylether propane, 1-butoxy-... (butoxymethyl)oxir... butyl glycidyl ether... oxirane, (butoxym...	methyl vinyl sulfone (methylsulphonyl)eth... (methylsulfonyl)eth... ethene, (methylsulf...	[(2-bromoethyl)thio... [(2-bromoethyl)sulf... 2-bromoethyl phen...	4-pentyn-1-ol pent-4-yn-1-ol	1-chloro-4-nitroben... chloro-4-nitrobenze... p-chloronitrobenzene benzene, 1-chloro-... 1chloro4nitrobenzene 4-chloronitrobenzene	4-nitrobenzyl bromide p-nitrobenzyl bromide 1-bromomethyl-4-n... 4-nitrobenzylbromide 1-(bromomethyl)-4-... benzene, 1-(bromo... ?-bromo-4-nitrotolu...	4-ethylnitroben... p-ethylnitroben... 1-ethyl-4-nitrc... benzene, 1-ethyl-... 4-ethyl-1-nitro...	EINECS:2028206
Chemical Name									
Molecular Formula	C8H9BrO	C7H14O2	C3H6O2S	C8H9BrS	C5H8O	C6H4ClNO2	C7H6BrNO2	C8H9NO2	
Structural Formula	Cc1ccc(Br)cc(C)c1O	CCCCOCC1CO1	CS(=O)(=O)C=C	BrCCSc1ccccc1	C#CCCCO	O=N(=O)c1ccc(Cl)...	O=N(=O)c1ccc(CBr...	CCc1ccc(N(=...	
Parameters									
Physical Chemical Properties									
Environmental Fate and Transport									
Ecotoxicological Information									
Human Health Hazards									
Predefined: Ecotoxicological Information									

1. Select the ECOSAR profiler by clicking on the box next to it (a green tick mark appears); 2. Click Apply.

Profiling

Profiling results

- 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 (see next screen shot).
- Please note the specific profiling results of Aquatic toxicity classification by ECOSAR.
- These results will be used to division chemicals from the inventory into suitable clusters in the next steps of the exercise.

Profiling

Profiling results

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below the menu is a toolbar with 'Apply', 'New', 'View', and 'Delete' buttons. The main window is divided into several panels. On the left, there are two panels for 'Profiling methods' and 'Metabolism/Transformations', both with 'Select All', 'Unselect All', and 'Invert' buttons. The central panel shows a 'Filter endpoint tree...' with a tree view containing 'Structure', 'Substance Identity', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Human Health Hazards', 'Profile', 'Endpoint Specific', and 'Aquatic toxicity classification by ECOSAR'. The 'ECOSAR' row is circled in red. To the right of the tree is a table with 8 columns labeled '1 [target]' through '8 [target]'. The table contains chemical structures in the first row and various data points in the second row. The 'ECOSAR' row contains the following categories: Phenols, Epoxides, mono, Vinyl/Allyl Sulfones, Neutral Organics, Neutral Organics, Neutral Organics, Benzyl Halides, and Neutri... The bottom status bar shows '300 Document' on the left and '1/0/0' on the right.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
 - Chemical Input
 - Profiling
 - **Endpoint**

Endpoints

Overview

- “Endpoint” refers to the electronic process of retrieving the environmental fate and toxicity data stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).

Endpoints

Case study

- In this example, since we categorize an inventory for later analysis we do not need of data gathering.
- In this module we need only to select the Aquatic ECETOC, Aquatic OASIS and Aquatic ECOTOX databases.
- These databases contain results on aquatic toxicity for various species, and durations.
- **Click** on the “Category Definition” to move to the next module.

Endpoints Case study

The screenshot shows the QSAR Toolbox interface with the following components:

- Callout 1:** Points to the 'Ecotoxicological Information' section in the left sidebar.
- Callout 2:** Points to the 'Aquatic ECOTOC', 'Aquatic Japan MoE', and 'Aquatic OASIS' checkboxes.
- Callout 3:** Points to the 'Gather' button in the top toolbar.

The main window displays a data matrix with the following structure:

Filter endpoint tree...	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]	8 [target]
Substance Identity								
Aquatic Toxicity (237/16920)	M: 13.6 mg/L	M: 3.9(3.1;4.9) mg...	M: 10.9 mg/L		M: 2.43E3 mg/L, 2...	M: 0.19 mg/L, 0.38...	M: 0.108 mg/L	M: 45.6 mg/L, 50.1...
Terrestrial Toxicity (103/3071)						M: 80 mg/L		

1. **Expand** Ecotoxicological information section 2. Select the databases listed in the previous slide both databases mentioned in the previous slide databases 2. **Click** Gather. The data appears on data matrix

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
 - Chemical Input
 - Profiling
 - Endpoint
 - **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 to define a category.

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.
- Detailed information about grouping chemical (Chapter 4) could be downloaded from:
<http://www.oecd.org/dataoecd/58/56/46210452.pdf>

Category Definition

Side bar to ECOSAR category

- ECOSAR has been used by the U.S. Environmental Protection Agency to predict the aquatic toxicity of new industrial chemicals in the absence of test data.
- SARs are developed for chemical classes (e.g. phenols) based on measured test data which have been submitted by industry or they were taken from other sources.
- The ECOSAR classification in the Toolbox is used for grouping chemicals by structural (and in some cases mechanistic) similarity.

Category Definition Clustering

The chemicals in the inventory can be clustered in to 58 clusters based on EcoSAR Categories.

- 67 are neutral organics
- 30 are esters
- 28 are phenols
- 23 are anilines (unhindered)
- 14 are aldehydes (mono)
- 10 are aliphatic amines
- Other clusters consist of less than 10 compounds

Category Definition Clustering

The screenshot shows the QSAR Toolbox software interface. The top toolbar includes buttons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The 'Category Definition' menu is open, showing options like Define, Define with metabolism, Subcategorize, Combine, Clustering, Delete, and Delete All. The 'Clustering' button is highlighted with a red circle and a '2' callout. On the left, the 'Endpoint Specific' list is visible, with 'Aquatic toxicity classification by ECOSAR' highlighted by a red circle and a '1' callout. The main window displays a 'Filter endpoint tree...' on the left and a table of chemical structures and their classifications on the right.

Structure	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]
<chem>Cc1ccc(O)c(C)c1</chem>	<chem>C1CCOC1</chem>	<chem>CC(C)S(=O)(=O)C</chem>	<chem>Cc1ccc(S)cc1</chem>	<chem>CC(C)CC</chem>	<chem>Cc1ccc(O)cc1</chem>	<chem>Cc1ccc(O)cc1</chem>
Substance Identity						
Parameters						
Physical Chemical Properties						
Environmental Fate and Transport						
Ecotoxicological Information (240/18379)	M: 13.6 mg/L	M: 22(18;28) mg/L,...	M: 10.9 mg/L		M: 2.43E3 mg/L, 2...	M: 0.19 mg/L
Human Health Hazards						
Profile						
Endpoint Specific						
Aquatic toxicity classification by ECO...	Phenols	Epoxides, mono	Vinyl/Allyl Sulfones	Neutral Organics	Neutral Organics	Neutral Organi

300 Document

1/0/0

1. Highlight the "Aquatic toxicity classification by ECOSAR";
2. Click **Clustering**.

Category Definition

Select working cluster

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' workflow selected. The 'Defined Categories' list on the left includes '28 STRICT(Phenols)', which is highlighted with a red circle and a callout box containing the number '1'. The main table displays chemical structures and their properties across eight target categories. The 'Endpoint Specific' category is highlighted in blue, and 'Phenols' is listed under this category for each target.

Structure	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]	8
Substance Identity								
Physical Chemical Properties								
Environmental Fate and Transport								
Ecotoxicological Information (28/5893)	M: 13.6 mg/L	M: 8.2 mg/L, 0.319...	M: 0.5 mg/L, 0.75 ...	M: 435 mg/L, 433(...	M: 18.9 mg/L, 7.1 ...	M: 6.6 mg/L, 6.6 ...	M: 3.54 mg/L, 3.46... M	
Human Health Hazards								
Profile								
Endpoint Specific								
Aquatic toxicity classification by EC...	Phenols	Phenols	Phenols	Phenols	Phenols	Phenols	Phenols	P

1. Highlight specific cluster to select it as working- in this example select "Phenols".

Category Definition

Select working cluster

The screenshot shows the QSAR Toolbox interface with the following components:

- Navigation Bar:** Input, Profiling, Endpoint, **Category Definition**, Data Gap Filling, Report.
- Workflow:** Categorize, Delete.
- Grouping Methods:**
 - Toxic hazard classification by Cramer (original)
 - Ultimate biodeg
 - Endpoint Specific**
 - Acute aquatic toxicity classification by Verhaar (Modified)
 - Acute aquatic toxicity MOA by OASIS
 - Aquatic toxicity classification by ECOSAR
 - Bioaccumulation - metabolism alerts
 - Bioaccumulation - metabolism half-lives
 - Biodegradation fragments (BioWIN MITI)
 - Carcinogenicity (genotox and nongenotox) alerts by ISS
 - DART scheme v.1.0
 - DNA alerts for Ames by OASIS v.1.4
 - DNA alerts for CA and MNT by OASIS v.1.1
 - Eye irritation/corrosion Exclusion rules by BfR
 - Eye irritation/corrosion Inclusion rules by BfR
 - in vitro mutagenicity (Ames test) alerts by ISS
 - in vivo mutagenicity (Micronucleus) alerts by ISS
 - Keratinocyte gene expression
 - Oncologic Primary Classification
 - Protein binding alerts for Chromosomal aberration by OASIS v.1.4
 - Protein binding alerts for skin sensitization by OASIS v1.4
 - Respiratory sensitisation
 - Retinoic Acid Receptor Binding
- Defined Categories:**
 - Document
 - 58] Clustering by Aquatic toxicity classification by ECOSAR
 - 67] STRICT(Neutral Organics)
 - 30] STRICT(Esters)
 - 28] STRICT(Phenols)
 - 23] STRICT(Anilines (Unhindered))
 - 14] STRICT(Aldehydes (Mono))
 - 11] STRICT(Not Related to an Existing ECOSAR Class)
 - 10] STRICT(Aliphatic Amines)
 - 8] STRICT(Propargyl Alcohols)
 - 7] STRICT(I/VA)
 - 6] STRICT(Polynitrobenzenes)
 - 5] STRICT(Anilines (Hindered))
 - 51] STRICT(Methacrylates)
- Main Table:**

Structure	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]
<i>Oncorhynchus tshawytscha</i> (1/2)							
<i>Oreochromis mossambicus</i> (1/3)			M: 19 mg/L, 24 mg...				
<i>Oryzias latipes</i> (3/4)			M: 29.3(23.4;36.6) ...				
<i>Oryzias melo</i> (1/2)			M: 9.3(8.3;10.4) m...				
<i>Perca fluviatilis</i> (1/1)							
<i>Phoxinus phoxinus</i> (1/1)					M: 9.5(6.9;13.2) mg/L		
<i>Pimephales promelas</i> (9/98)		M: 16.4 mg/L, 28.6...		M: 32(23.8;39.9) m...		M: 6.6 mg/L, 6.8(4...	M: 3.46 m
<i>Platichthys flesus</i> (3/3)				M: 20 mg/L, 20.9 ...			
<i>Pleuronectes platessa</i> (1/5)							
<i>Poecilia reticulata</i> (6/27)		M: 19.7 mg/L		M: 39.2(34.1;47.6) ...		M: 6.6 mg/L	
<i>Poeciliopsis occidentalis</i> (1/3)							
<i>Pomatoschistus microps</i> (1/1)							
<i>Ptychocheilus lucius</i> (1/6)							
<i>Puntius sophore</i> (1/2)				M: 14.5(10.8;19.4) ...			
<i>Rasbora daniconius neilgerie...</i> (1/1)							
<i>Rasbora heteromorpha</i> (1/1)							
<i>Rutilus rutilus</i> (1/1)							
<i>Salmo salar</i> (1/19)							
<i>Salmo trutta ssp. lacustris</i> (1/1)							
<i>Salvelinus malma</i> (1/1)							
<i>Solea solea</i> (2/2)				M: 14.2 mg/L			
<i>Thymallus arcticus</i> (1/1)							
<i>Umbra pygmaea</i> (1/1)							

Category Definition

Navigation through the endpoint tree

- The user can navigate through the data tree by closing or opening the nodes of the tree.
- In this example, the LC50-96h Mortality for *Pimephales promelas* is the target endpoint.

Category Definition

Navigation through the endpoint tree-target

The screenshot shows the QSAR Toolbox interface with the 'Category Definition' workflow selected. The 'Endpoint' tab is active, displaying a list of species and their associated toxicity data across seven target categories. The row for 'Pimephales promelas' is highlighted with a red circle.

Species	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]
<i>Oncorhynchus tshawytscha</i> (1/2)							
<i>Oreochromis mossambicus</i> (1/3)			M: 19 mg/L, 24 mg...				
<i>Oryzias latipes</i> (3/4)			M: 29.3(23.4;36.6) ...				
<i>Oryzias melastigma</i> (1/2)			M: 9.3(8.3;10.4) m...				
<i>Perca fluviatilis</i>							
<i>Phoxinus phoxinus</i> (1/1)			M: 9.5(6.9;12.9) mg/L				
<i>Pimephales promelas</i> (9/98)		M: 16.4 mg/L, 28.6...	M: 32(23.8;39.9) m...			M: 6.6 mg/L, 6.8(4...	M: 3.1...
<i>Platichthys flesus</i> (3/5)			M: 20 mg/L, 20.8 m...				
<i>Pleuronectes platessa</i> (1/5)							
<i>Poecilia reticulata</i> (6/27)		M: 19.7 mg/L	M: 39.2(34.1;47.6) ...			M: 6.6 mg/L	
<i>Poeciliopsis occidentalis</i> (1/3)							
<i>Pomatoschistus microps</i>							
<i>Ptychocheilus lucius</i> (1/6)							
<i>Puntius sophore</i> (1/2)			M: 14.5(10.8;19.4) ...				
<i>Rasbora daniconius neilgerie...</i> (1/1)							
<i>Rasbora heteromorpha</i>							
<i>Rutilus rutilus</i> (1/1)							
<i>Salmo salar</i> (1/19)							
<i>Salmo trutta ssp. lacustris</i> (1/1)							
<i>Salvelinus malma</i>							
<i>Solea solea</i> (2/2)		M: 14.2 mg/L					
<i>Terapon jarbua</i> (1/1)		M: 40 mg/L					
<i>Thymallus arcticus</i>							
<i>Tilapia zillii</i>							
<i>Umbra pygmaea</i>							

The Exercise - Part I

Recap

- We have loaded in the Toolbox data matrix a user predefined list of 300 chemicals.
- The inventory has been profiled by the “Aquatic toxicity classification by ECOSAR” scheme.
- The inventory has been clustered by the ECOSAR categorization method.
- The “Phenols” cluster has been selected as the working cluster and experimental data have been gathered for all its members.
- Now we are ready to proceed with the Part II of the exercise.

Outlook

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

Data gap filling

Overview

- “Data Gap Filling” module give access to three different data gap filling tools:
 - Read-across
 - Trend analysis
 - (Q)SAR models
- Select the most relevant data gap method:
 - Read-across is the appropriate data-gap filling method for “qualitative” endpoints where possible results are positive, negative or equivocal. Read-across is also recommended for “quantitative endpoints” (e.g., 96h-LC50 for fish) when there are few analogues with experimental results.
 - Trend analysis is the appropriate data-gap filling method for “quantitative endpoints” (e.g., 96h-LC50 for fish) when a high number of analogues with experimental results.
 - “(Q)SAR models” can be used to fill a data gap if no adequate analogues are found for a target chemical.
- In this example we will use trend analysis.

Data gap filling

Apply Trend analysis

The screenshot shows the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' method is selected in the sidebar, with 'Trend analysis' chosen. The main table displays data for various target endpoints. The row for 'Pimephales promelas' under target endpoint 4 is highlighted. A callout box at the bottom provides instructions: 1. Highlight the data endpoint box corresponding to the target endpoint under 4-bromo-2,6-dimethylphenol; 2. Select Trend analysis; 3. Click Apply.

Structure	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]	8 [target]
Oryzias latipes (3/4)			M: 29.3(23.4;36.6) ...					
Oryzias melastigma (1/2)			M: 9.3(8.3;10.4) m...					
Perca fluviatilis								
Phoxinus phoxinus (1/1)			M: 9.5(6.9;13.2) mg/L					
Pimephales promelas (9/98)	M: 16.4 mg/L, 28.6...		M: 32(23.8;39.9) m...			M: 6.6 mg/L, 6.8(4...		M: 3.46 mg/L, 3.46...
Platichthys flesus (3/5)			M: 20 mg/L, 20.9 ...					
Pleuronectes platessa (1/5)								
Poecilia reticulata (6/27)	M: 19.7 mg/L		M: 39.2(34.1;47.6) ...			M: 6.6 mg/L		
Poeciliopsis occidentalis (1/3)								
Pomatoschistus microps								
Ptychocheilus lucius (1/6)								
Puntius sophore (1/2)			M: 14.5(10.8;19.4) ...					
Rasbora daniconius neilger... (1/1)								
Rasbora heteromorpha								
Rutilus rutilus (1/1)								
Salmo salar (1/19)								
Salmo trutta ssp. lacustris (1/1)								
Salvelinus malma								
Solea solea (2/2)			M: 14.2 mg/L					
Terapon jarbua (1/1)			M: 40 mg/L					
Thymallus arcticus								
Tilapia zillii								

28 STRICT(Phenols)

1. **Highlight** the data endpoint box corresponding to the target endpoint under 4-bromo-2,6-dimethylphenol; 2. **Select** Trend analysis; 3. **Click** Apply.

Data gap filling

Trend analysis

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Filling

Apply

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models

Target Endpoint

Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animals Chordata Actinopterygii Pimephales pimephales

Structure

1 [target] 2 [target] 3 [target] 6 [target] 7 [target] 20 [target] 22 [target] 23 [target]

Structure

Pimephales pimephales (9/98)

M: 16.4 mg/L, 28.6... M: 32(23.8, 39.9) m... M: 6.6 mg/L, 6.8(4... M: 3.46 mg/L, 3.46... M: 0.0986(0.0806;... M: 4.66 mg/L, 4.24... M: 6.18 mg/L

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 9 values from 9 analogue chemicals, Observed target value: N/A, Predicted target value: 3.58 mg/L, Model equation: $LC50 = +2.37 + 0.680 * \log Kow$

Descriptor X: log Kow

Accept prediction

Return to matrix

- Select/filter data
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

28: STRICT(Phenols) Create prediction by gap filling 0/1 1/1/0

Data gap filling

Interpreting Trend analysis

- The resulting plot outlines the experimental LC50 results of all analogues (Y axis) according to a descriptor (X axis) with Log Kow being the default descriptor.
- The **RED** dot represents the estimated result for the target chemical.
- The **BLUE** dots represent the experimental results available for the analogues.

Data gap filling

Analysis of the results

- The resulting trend analysis reveals a consistent trend in results with toxicity increase linearly with log Kow.
- It can be used to fill the data gap.
- By extension it can be used to fill data gaps for this endpoint for all chemicals in this cluster (i.e., category).
- See also the presentation “How to build user defined QSAR model” available at the OECD/LMC websites.
- Accept the prediction (see next screen shot).

Data gap filling

Accepting the prediction result

The screenshot shows the QSAR Toolbox interface during a data gap filling process. The main workspace displays a table of chemical structures and their predicted LC50 values. The table has columns for target values and predicted values. A callout box '1' points to the 'Accept prediction' button in the right sidebar. Another callout box '2' points to the 'Return to matrix' button. Below the table, a trend analysis plot is shown, with a red regression line. The plot title is 'Trend analysis prediction of LC50, making a linear approximation, based on 9 values from 9 analogue chemicals, Observed target value: N/A, Predicted target value: 3.58 mg/L, Model equation: LC50 = +2.37 +0.680 * log Kow'. The plot shows log Kow on the x-axis and LC50 (obs., log10(mol/L)) on the y-axis.

1. Click Accept prediction and after that 2. Return to the data matrix

Data gap filling

Accepting the prediction result

QSAR TOOLBOX

Input Profiling Endpoint Category Definition **Data Gap Filling** Report

Filter endpoint tree... Structure

Target Endpoint

Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata Actinopterygii Perca fluviatilis

Read-across Trend analysis (Q)SAR models

	1 [target]	2 [target]	3 [target]	4 [target]	5 [target]	6 [target]	7 [target]	8 [target]
— <i>Oryzias latipes</i> (3/4)			M: 29.3(23.4;36.6) ...					
— <i>Oryzias melastigma</i> (1/2)			M: 9.3(8.3;10.4) m...					
— <i>Perca fluviatilis</i>								
— <i>Phoxinus phoxinus</i> (1/1)			M: 9.5(6.9;13.2) mg/L					
— <i>Pimephales promelas</i> (10/99)	T: 3.58(0.397;32.3)...	M: 16.4 mg/L, 28.6...	M: 32(23.8;39.9) m...			M: 6.6 mg/L, 6.8(4...		M: 3.46 mg/L, 3.46...
— <i>Platichthys flesus</i> (3/5)			M: 20 mg/L, 20.9 ...					
— <i>Pleuronectes platessa</i> (1/5)								
— <i>Poecilia reticulata</i> (6/27)		M: 19.7 mg/L	M: 39.2(34.1;47.6) ...			M: 6.6 mg/L		
— <i>Poeciliopsis occidentalis</i> (1/3)								
— <i>Pomatoschistus microps</i>								
— <i>Ptychocheilus lucius</i> (1/6)								
— <i>Puntius sophore</i> (1/2)			M: 14.5(10.8;19.4) ...					
— <i>Rasbora daniconius neilger...</i> (1/1)								
— <i>Rasbora heteromorpha</i>								
— <i>Rutilus rutilus</i> (1/1)								
— <i>Salmo salar</i> (1/19)								
— <i>Salmo trutta ssp. lacustris</i> (1/1)								
— <i>Salvelinus malma</i>								
— <i>Solea solea</i> (2/2)			M: 14.2 mg/L					
— <i>Terapon jarbua</i> (1/1)			M: 40 mg/L					
— <i>Thymallus arcticus</i>								
— <i>Tilapia zillii</i>								
— <i>Umbra pygmaea</i>								
— <i>Xyrauchen texanus</i> (1/5)								
— Undefined Test organisms (spe...								

28 STRICT(Phenols) 1/0/0

Filled data gap

- By accepting the prediction the data gap was filled (see next screen shot).
- You are now ready to complete the final module and to download the report.
- **Click** on “Report” to proceed to next module.

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- **Workflow**
 - Chemical Input
 - Profiling
 - Endpoint
 - Category definition
 - Data Gap Filling
- **Report**

Report Overview

- The Report module can generate a report on any of predictions performed with the Toolbox.
- This module contains predefined report templates as well as a template editor which allows for development of user defined templates.
- The report can be printed or saved in different formats.

Report Generate report

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The left sidebar shows a tree view with 'Predictions' selected. The main window displays a report titled 'Prediction of LC50 for 4-bromo-2,6-dimethylphenol'. A blue callout box at the bottom contains instructions: '1. Double click on "Prediction" in the window Available data to report; 2. Select the prediction for the target chemical; 3. Click Create.'

Outlook

- Background
- Objectives
- Specific Aims
- Trend analysis
- The exercise
- Workflow
- **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

The screenshot displays the QSAR Toolbox interface. The main window shows a prediction result for LC50 of 4-bromo-2,6-dimethylphenol. A 'Save As' dialog box is open, allowing the user to save the prediction result as a file. The dialog shows the file name 'Tutonal 9.tbw' and the save type 'Toolbox work file (*.tbw)'. Red callout boxes with numbers 1, 2, and 3 indicate the steps: 1. Click on Save button; 2. Define name of the file; 3. Click Save button.

1. Click on Save button; 2. Define name of the file; 3. Click Save button

Open saved file

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Input' menu is open, and the 'Open' option is highlighted. A file explorer window is open, showing a folder named 'New folder' containing a file named 'Tutorial 9.tbw'. The file is selected, and the 'Open' button is highlighted. The main window displays a chemical structure and a list of endpoints.

Once the file has been saved **1. Go** to Input; **2. Click** Open; **3. Find** and **select** file; **4. Click** Open

Congratulation

- You have used the Toolbox to categorize an inventory by mechanistic behavior.
- You now know another useful tool in the Toolbox.
- Continue to practice with this and other tool and soon you will be comfortable dealing with many situations where the Toolbox is useful.