



# QSAR Application Toolbox

## Workflow

Laboratory of Mathematical Chemistry,  
Bourgas University “Prof. Assen Zlatarov”  
Bulgaria

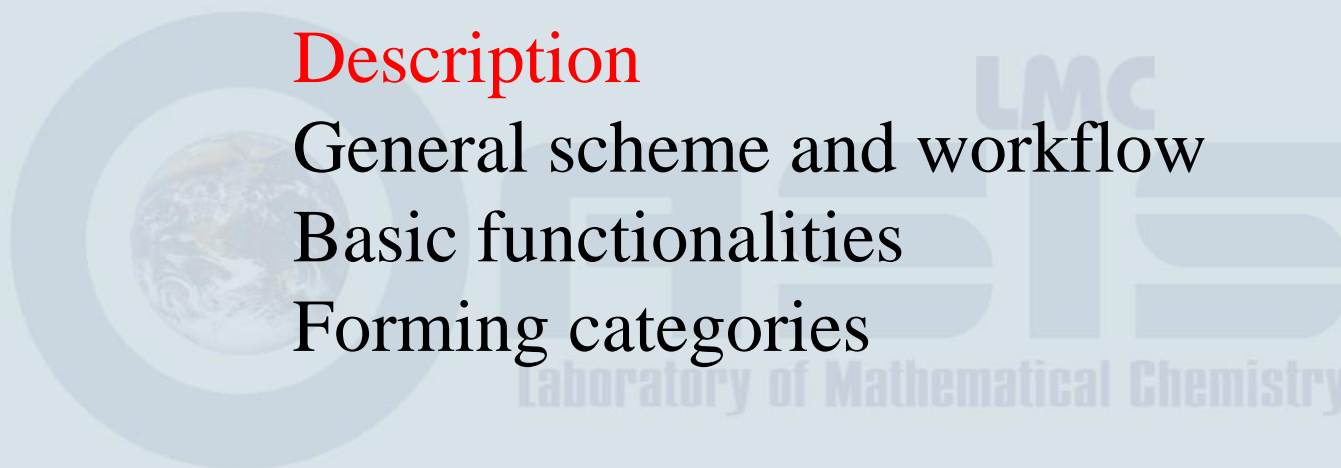
# Outlook

## Description

General scheme and workflow

Basic functionalities

Forming categories



# Philosophy

## QSAR TOOLBOX

- Toolbox helps registrants and authorities to
  - ✓ Use the methodologies to group chemicals into categories and
  - ✓ Refine and expand the categories approach
  - ✓ Provide a mechanistic transparency of the formed categories
  - ✓ Fill data gaps by read-across, trend analysis and (Q)SARs
  - ✓ Ensure uniform application of read-across
  - ✓ Support the regulatory use of (Q)SAR approach
  - ✓ Improve the regulatory acceptance of (Q)SAR methods

# Philosophy

QSAR TOOLBOX

- (Q)SAR Toolbox is a central tool for non-test data in ECHA and OECD
- ECHA and OECD coordinate the Toolbox development

# Description

## QSAR TOOLBOX

- Toolbox is not a QSAR model and hence can not be compared with other QSAR models
- Training sets (categories) for each prediction are defined dynamically as compared to other (Q)SAR model which have rigid training sets
- Each estimated value can be individually justified based on:
  - ✓ Category hypothesis (justification) and consistency
  - ✓ Quality of measured data and
  - ✓ Computational method used for grouping and data gap filling

# Description

## QSAR TOOLBOX

- It is developed under the continuing peer review of:
  - ✓ Member state countries,
  - ✓ Regulatory agencies
  - ✓ Chemical industry and NGOs
- The predictions are getting acceptance by toxicologists and regulators due to the:
  - ✓ International peer review process for developing the system and
  - ✓ Mechanistic transparency of the results
- The system is freely available and maintained in the public domain by OECD

# History

## **Phase I - The first version (2005 - 2008)**

- Emphasizes technological proof-of-concept
- Released in 2008
- Developed by LMC

## **Phase II (2008 - 2012)**

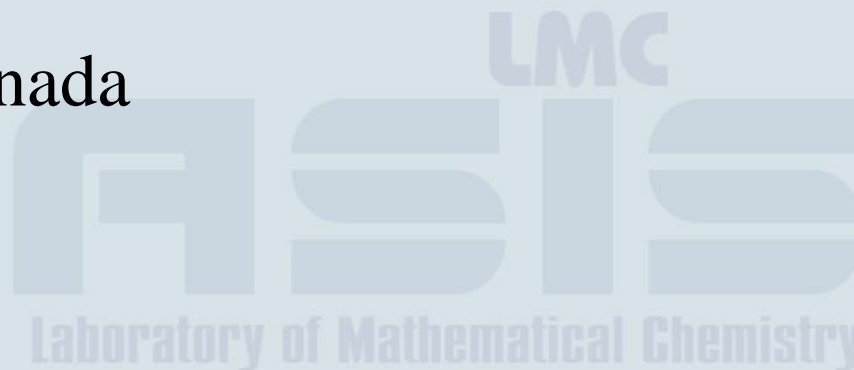
- More comprehensive Toolbox which fully implements the capabilities of the first version
- Second version was released in 2010
- Developed by LMC, subcontractors: LJMU, Lhasa Ltd and TNO
- Third version was released in 2012
- Maintenance 2013 - v. 3.2 – in January 2014
- Version 3.3.1 was released in December 2014
- Version 3.3.2 was released in February 2015
- Version 3.3.3 is expected to be released in June 30, 2015

## **Phase III (2013 – 2019 + 4 years maintenance)**

- Significant focus on streamlining + new IT platform
- Knowledge and data rationalization and curation
- Implementation of Ontology, ADME
- Implementation of AOPs
- Version 4.0 will be released in spring of 2016
- Developed by LMC, subcontractors: LJMU and Lhasa Ltd

# Main Government Donators

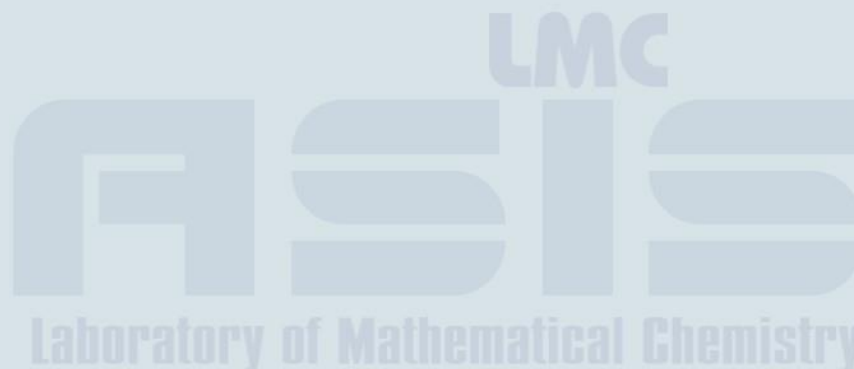
- OECD
- European chemicals agency
- US EPA
- Environment Canada
- Health Canada
- NITE Japan
- NIES Japan
- Danish EPA
- UBA Germany
- NICNAS Australia
- DEWNA Australia
- ISS Italy





# Main Industry Donators

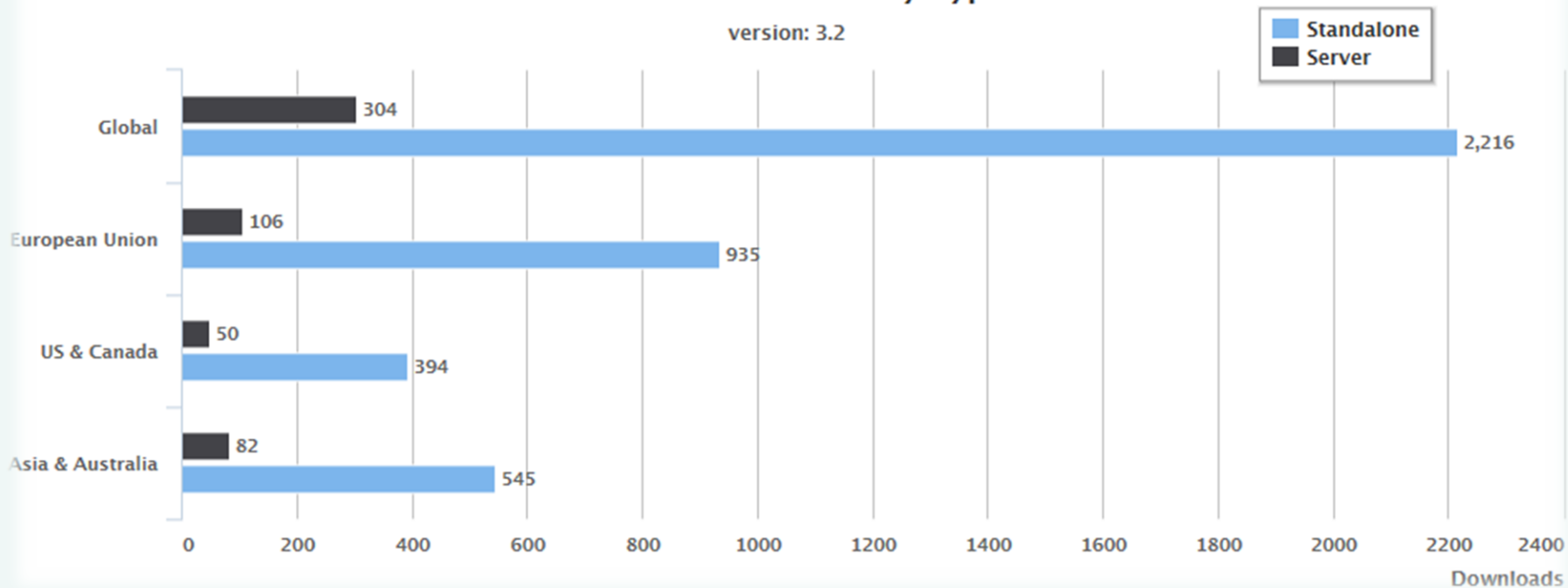
- L'Oreal
- DuPont
- Givaudan
- Dow chemicals
- BASF
- ExxonMobil
- 3M
- Firmenich SV
- SRC, Syracuse
- Unilever



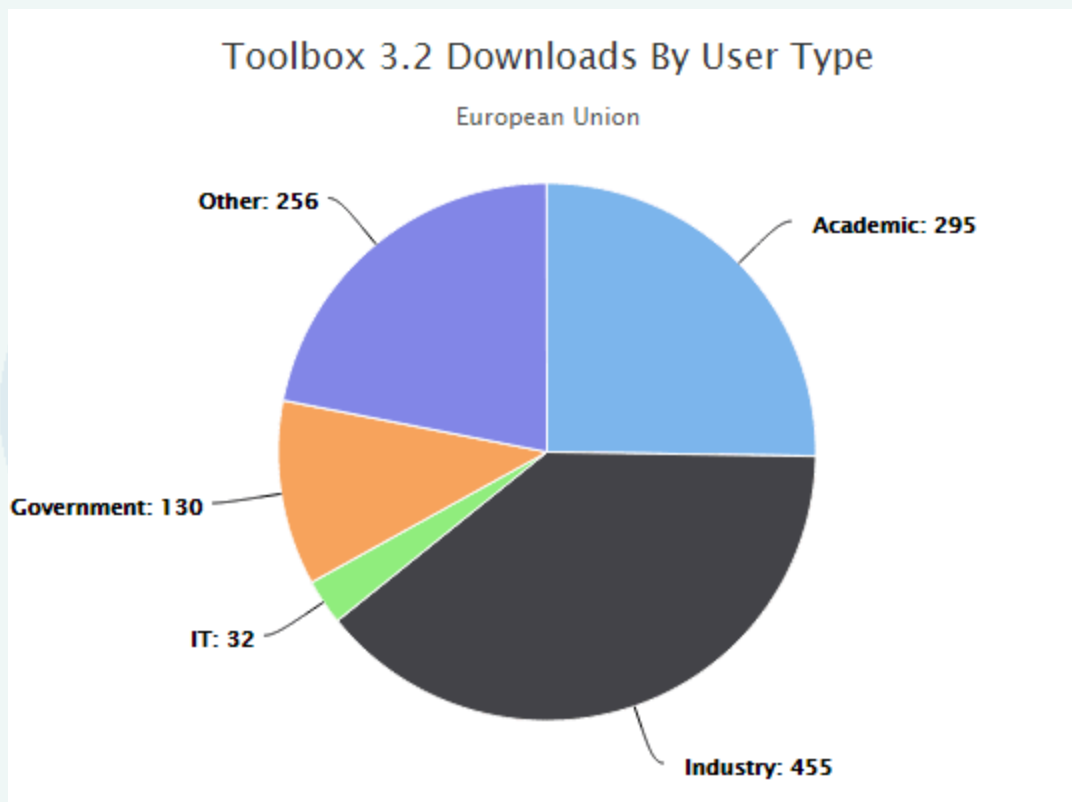
# Download statistics of QSAR Toolbox by registered users worldwide

## Toolbox Downloads By Type

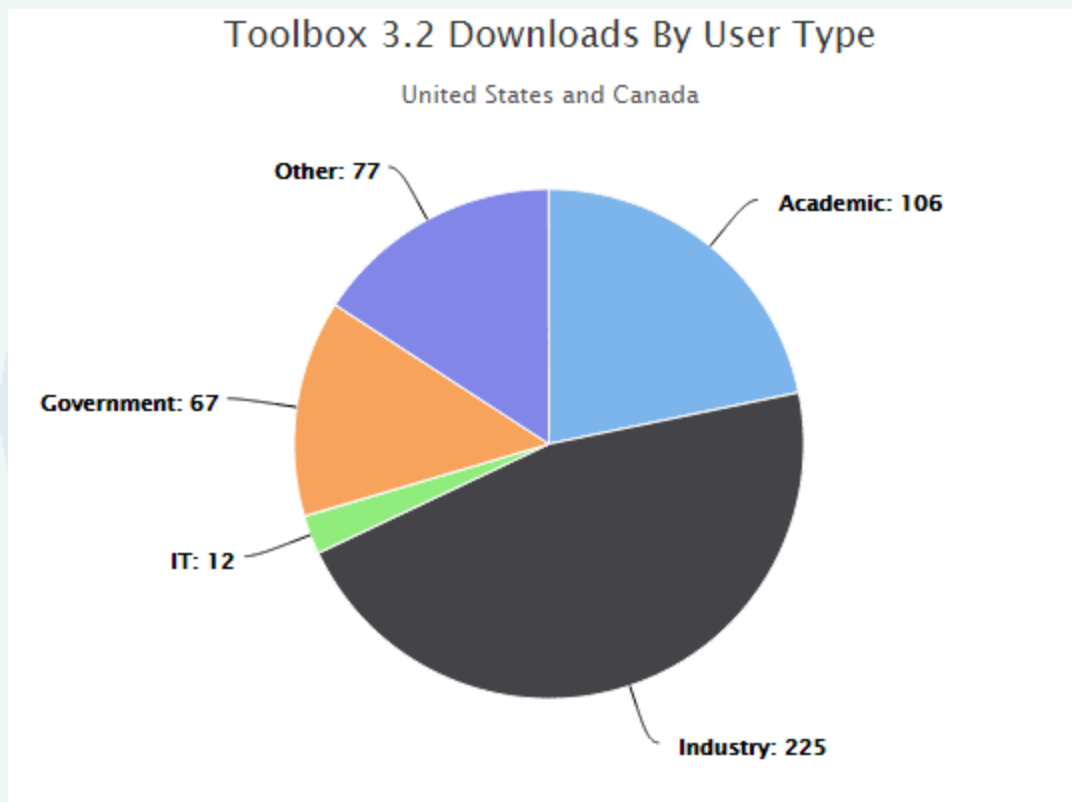
version: 3.2



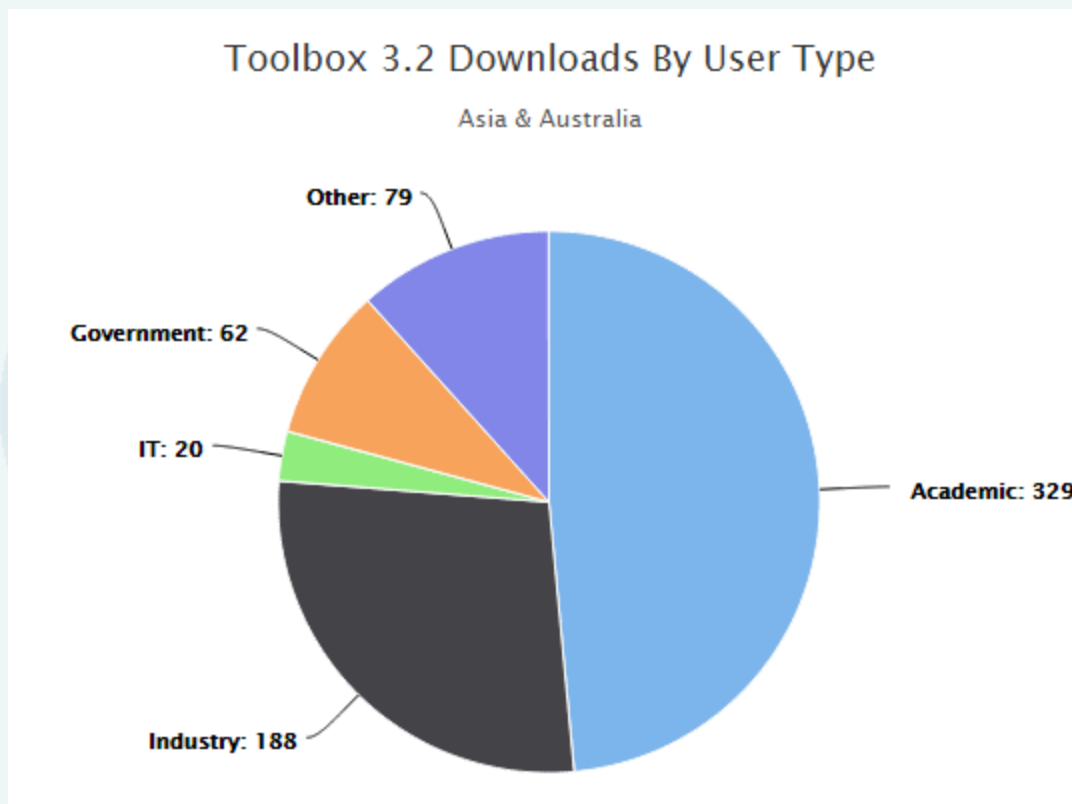
# Download statistics of QSAR Toolbox by registered users worldwide



# Download statistics of QSAR Toolbox by registered users worldwide

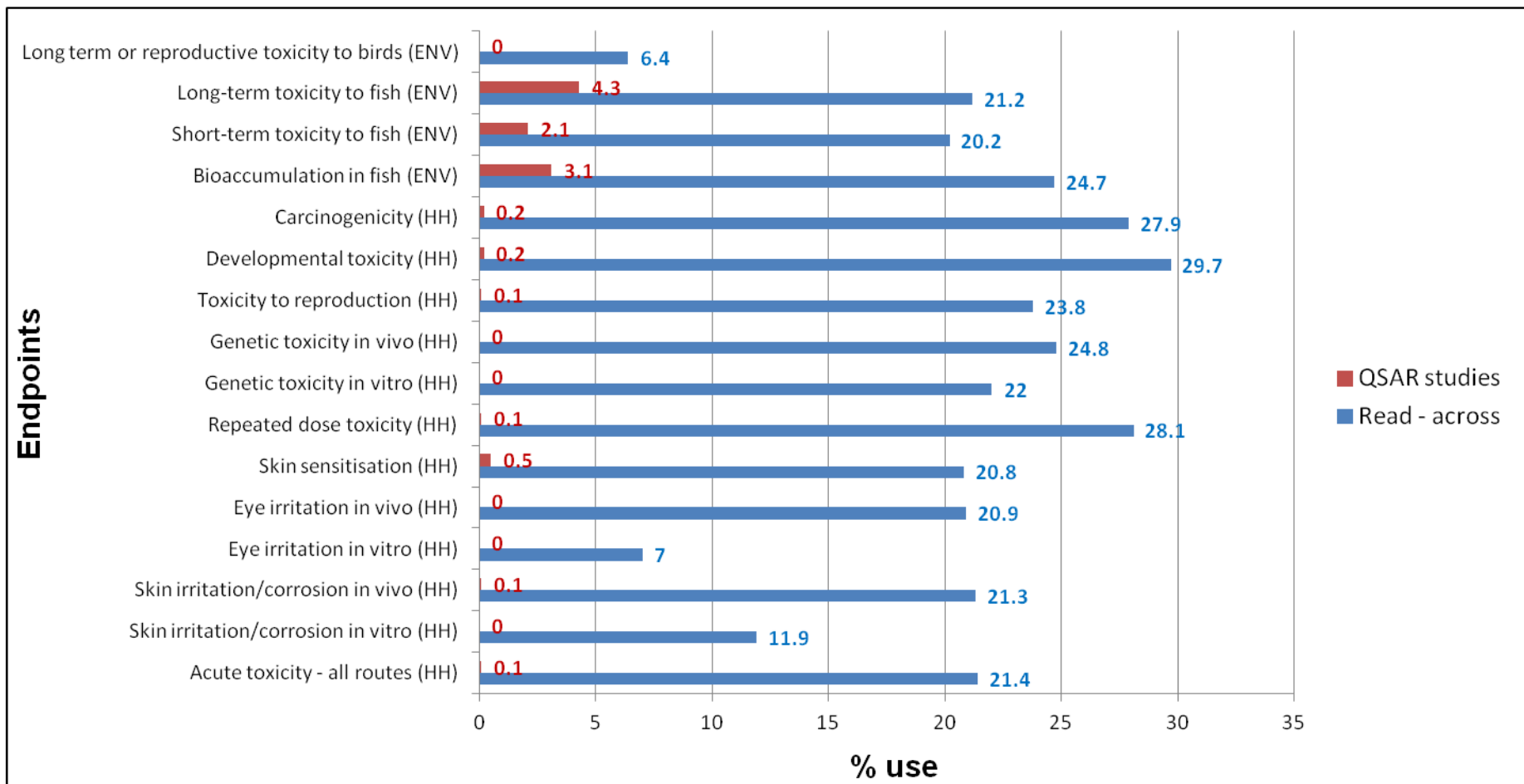


# Download statistics of QSAR Toolbox by registered users worldwide



# QSAR and read-across based submissions to the ECHA

For existing substances at or above 1000 tpa\*  
(2011 ECHA report)



ENV = environmental endpoint; HH = human health endpoint

\*H. Spielmann et al. ATLA 39, 481–493, 2011

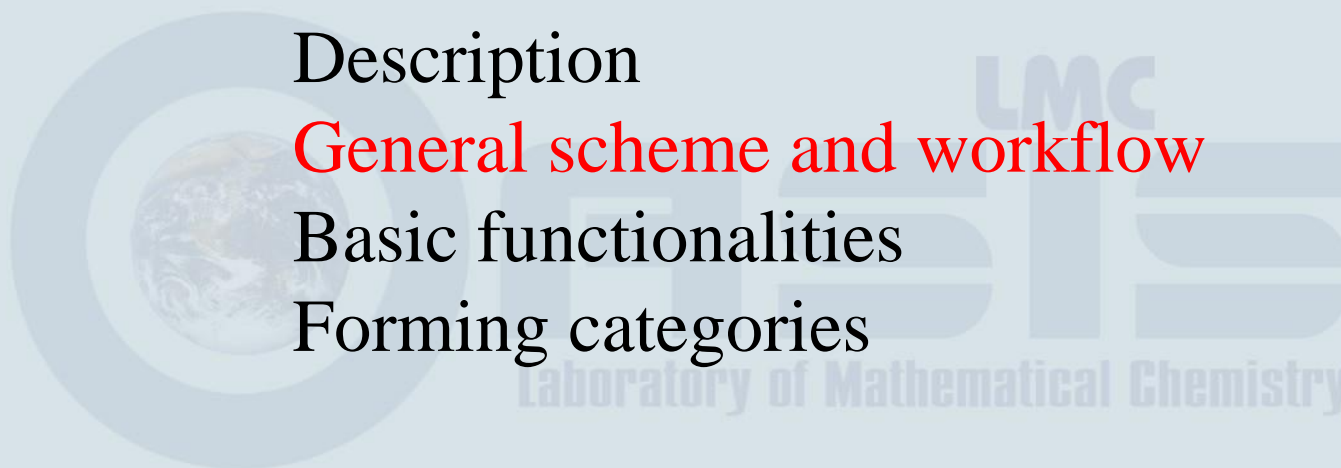
# Outlook

Description

General scheme and workflow

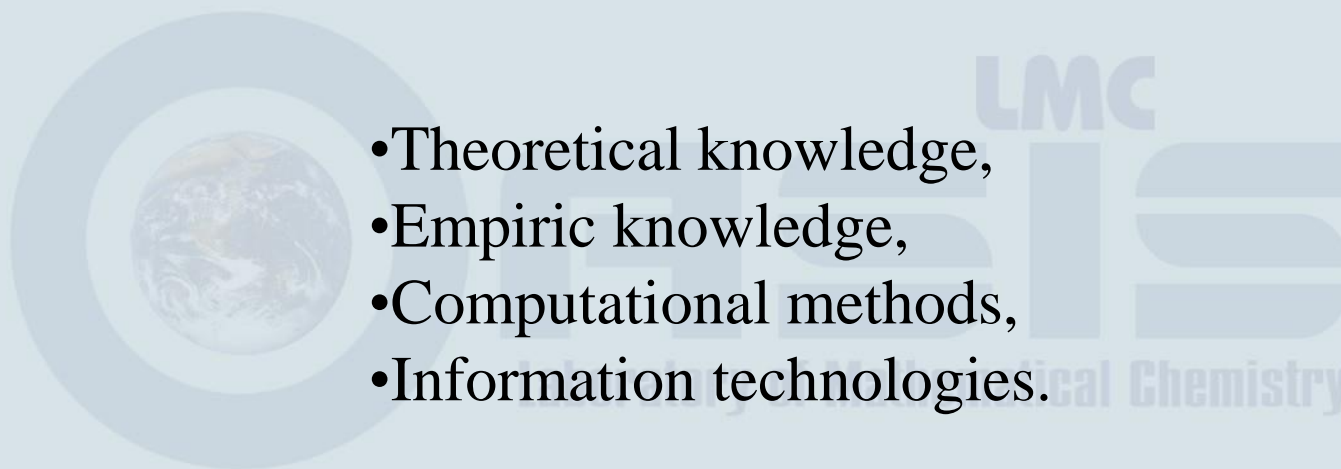
Basic functionalities

Forming categories



# General Scheme

- Theoretical knowledge,
- Empiric knowledge,
- Computational methods,
- Information technologies.

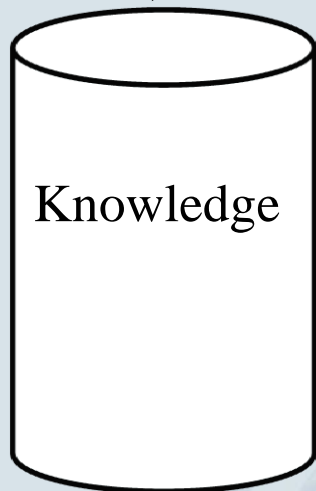




↓ Input

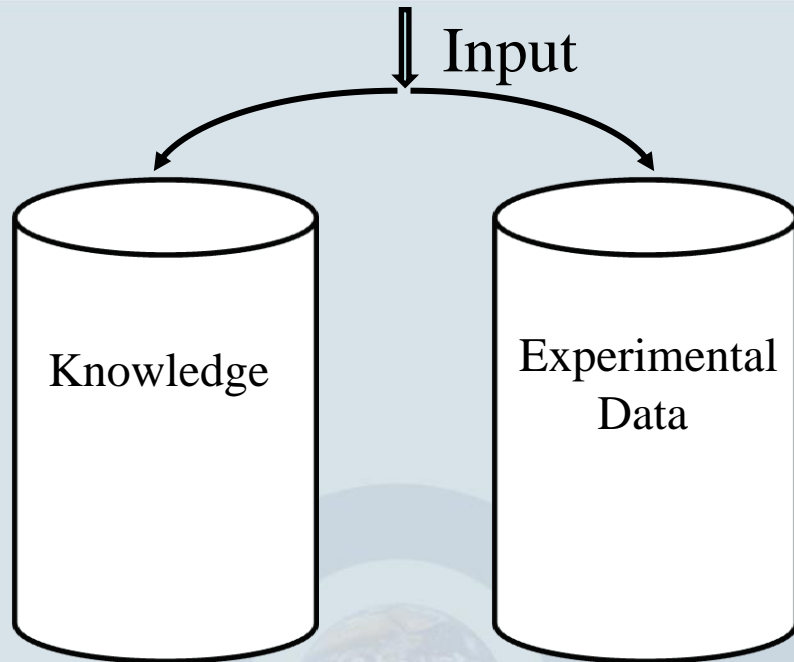


Input

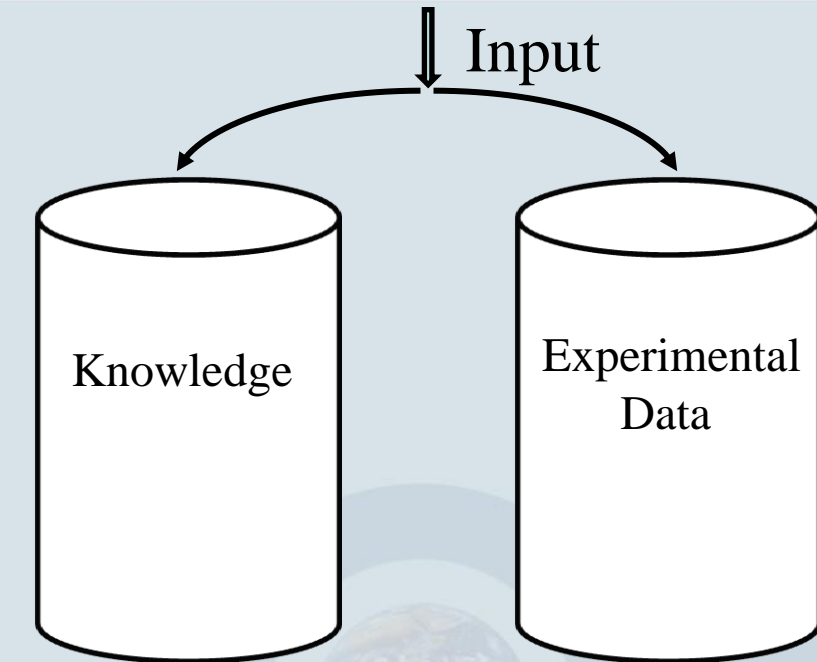


- Is the chemical included in regulatory or chemical categories?
- Could this chemical interact with DNA or specific proteins?
- Could this chemical cause adverse effects?
- Is the effect due to parent chemical or its metabolites?

LMC  
ASIS  
Laboratory of Mathematical Chemistry

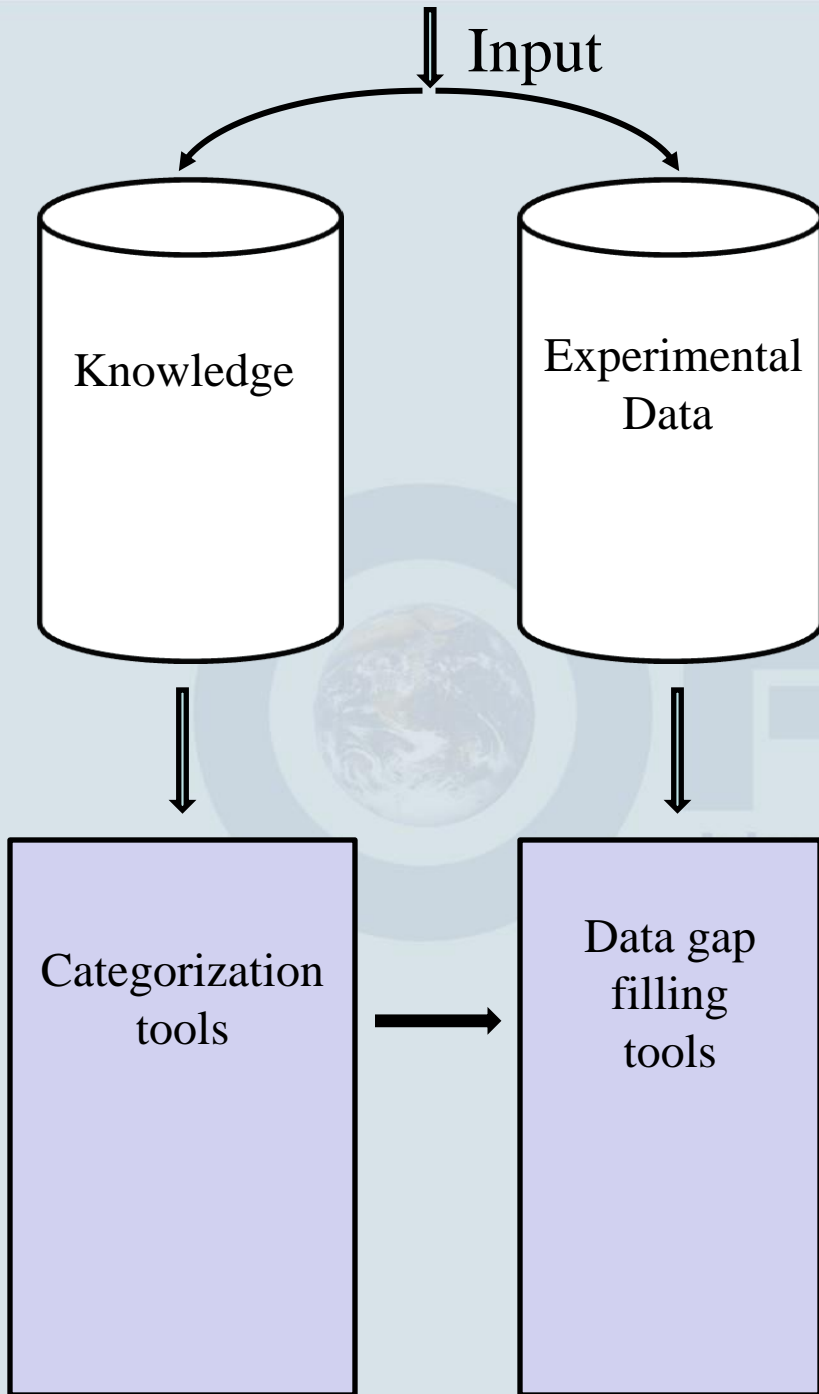


- Are data available for assessed endpoints of target chemical?
- Is information for the data sources available?

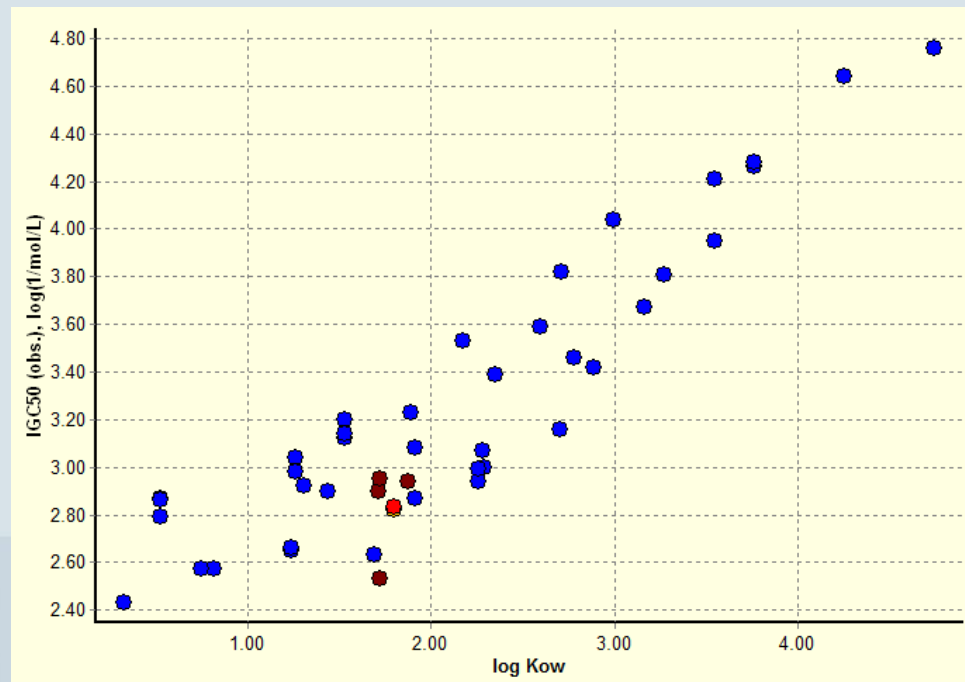
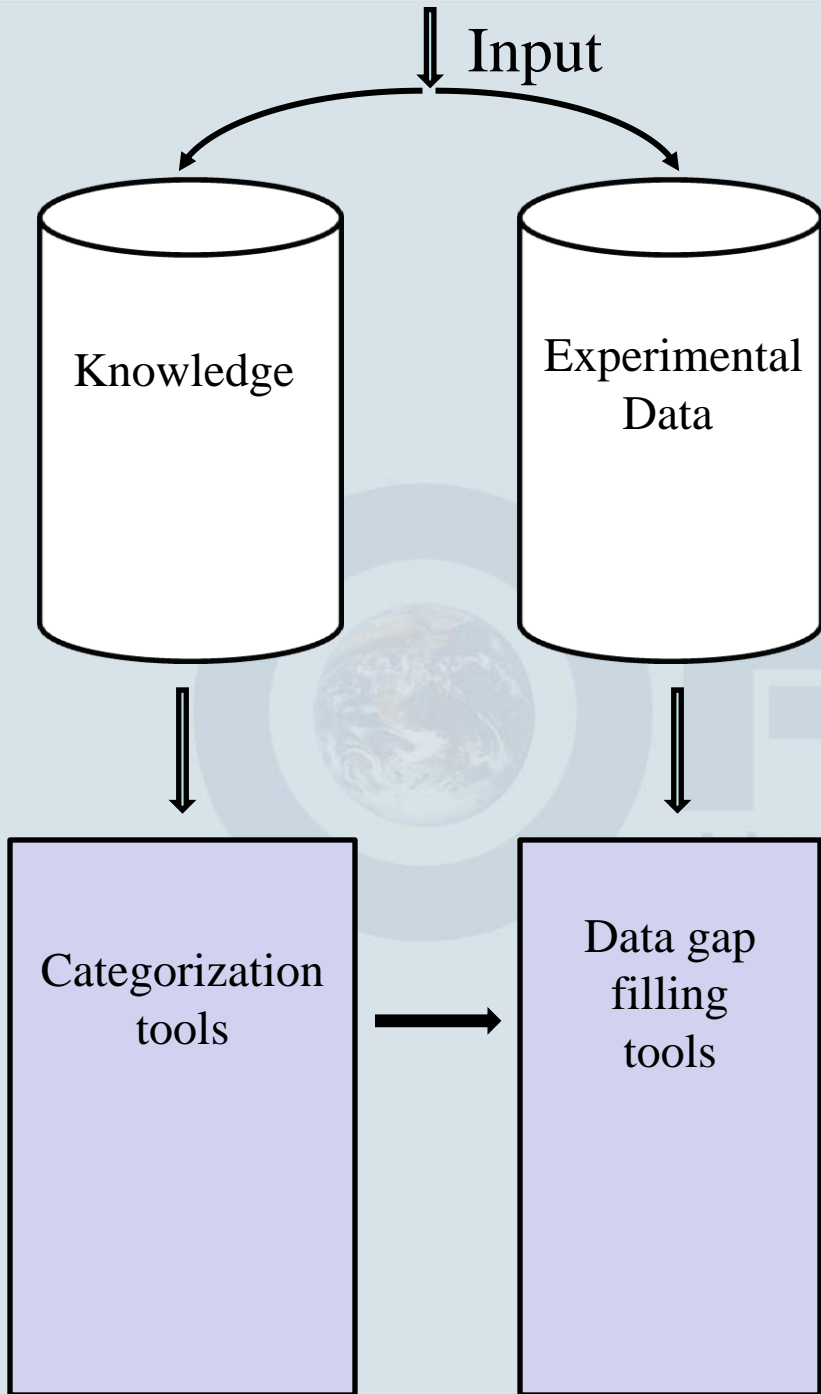


- Search for possible analogues using existing categorization schemes
- Group chemicals based on common chemical/toxicological mechanisms and/or metabolism
- Design a data matrix of a chemical category
- Set the endpoints hierarchy in the data matrix

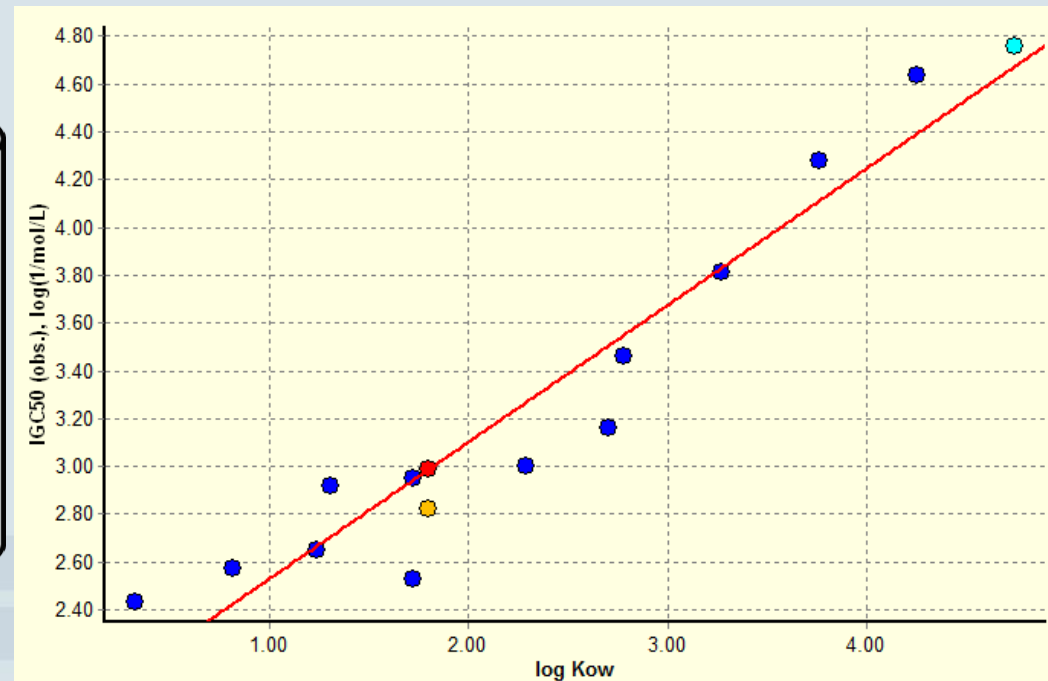
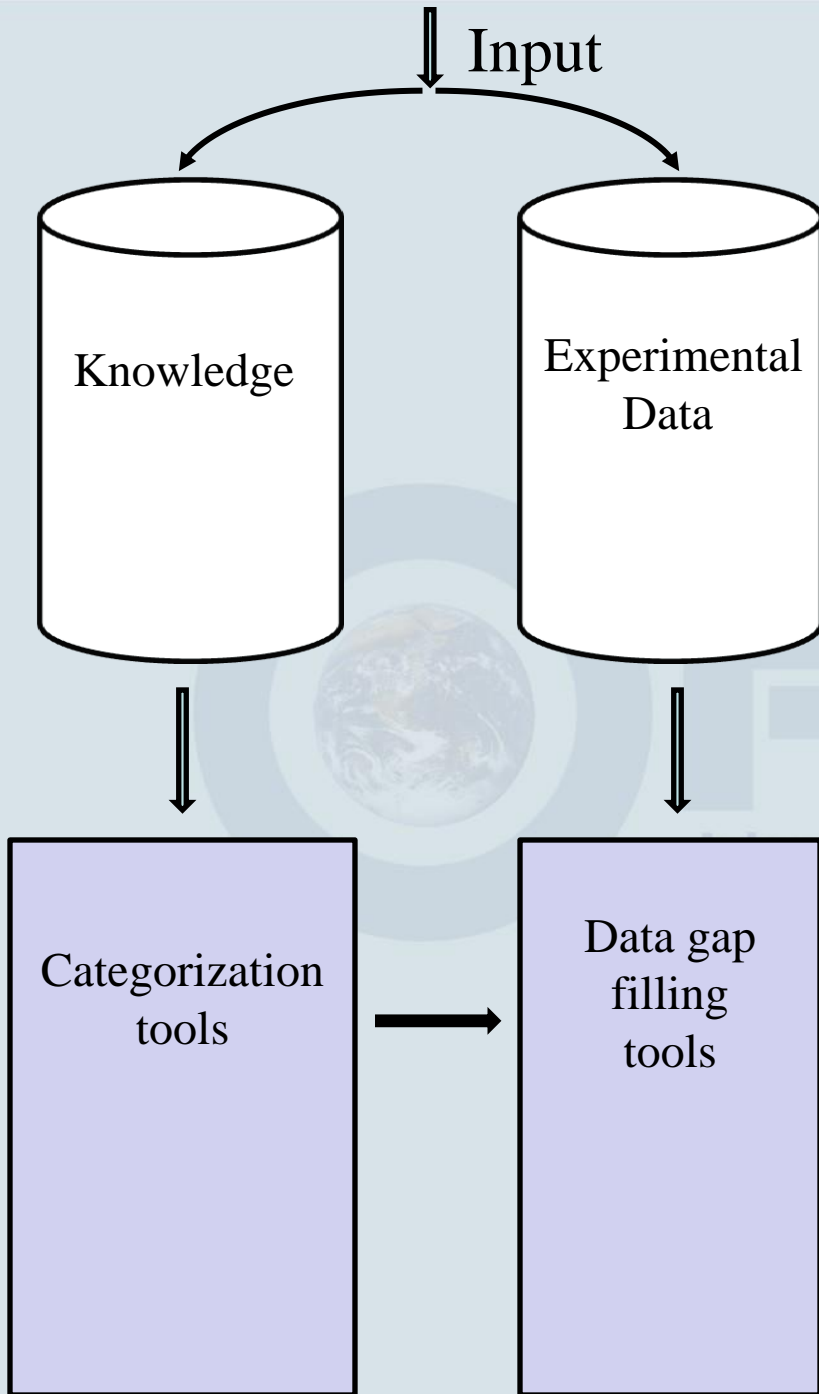
LMC  
Laboratory of Mathematical Chemistry



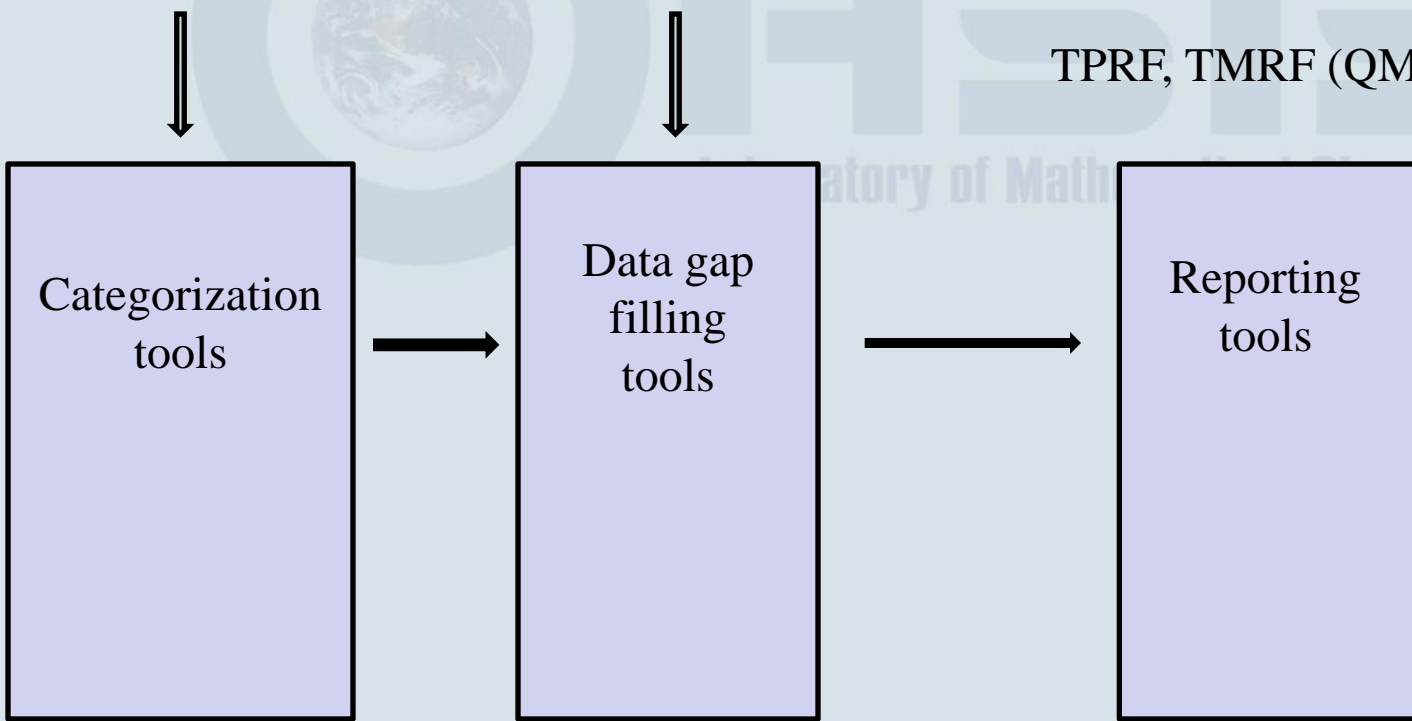
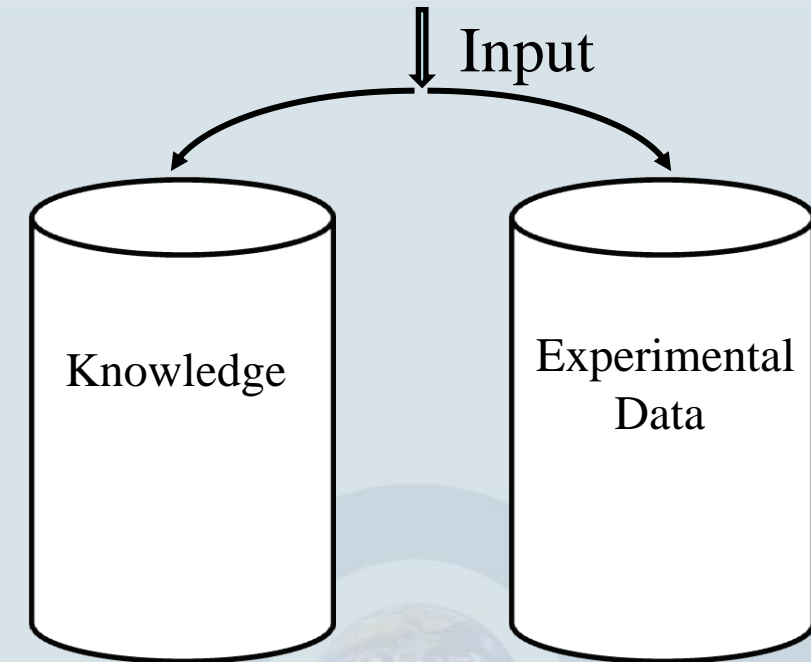
- Read-across
- Trend Analysis
- (Q)SAR



- **Read-across**
- Trend Analysis
- (Q)SAR

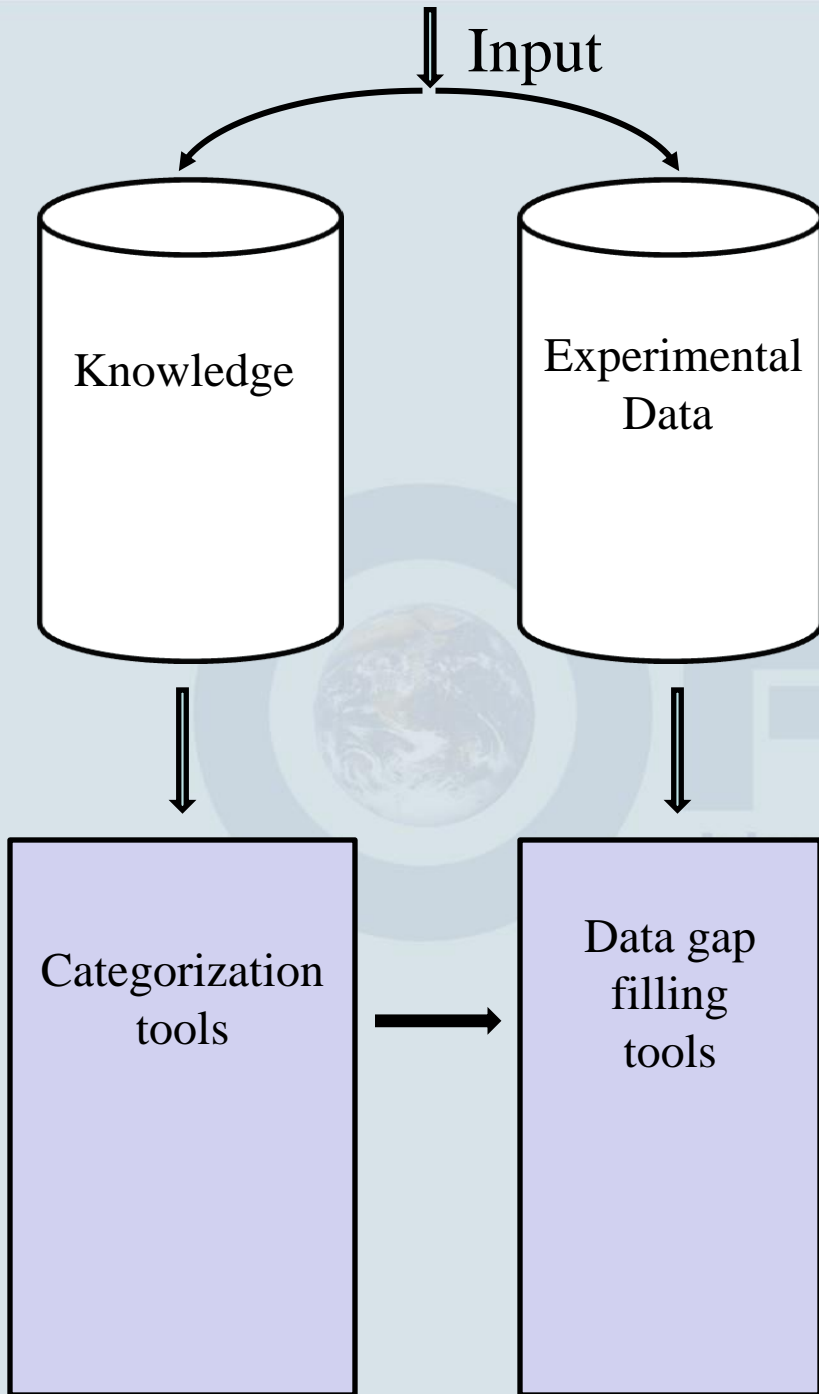


- Read-across
- **Trend Analysis**
- (Q)SAR

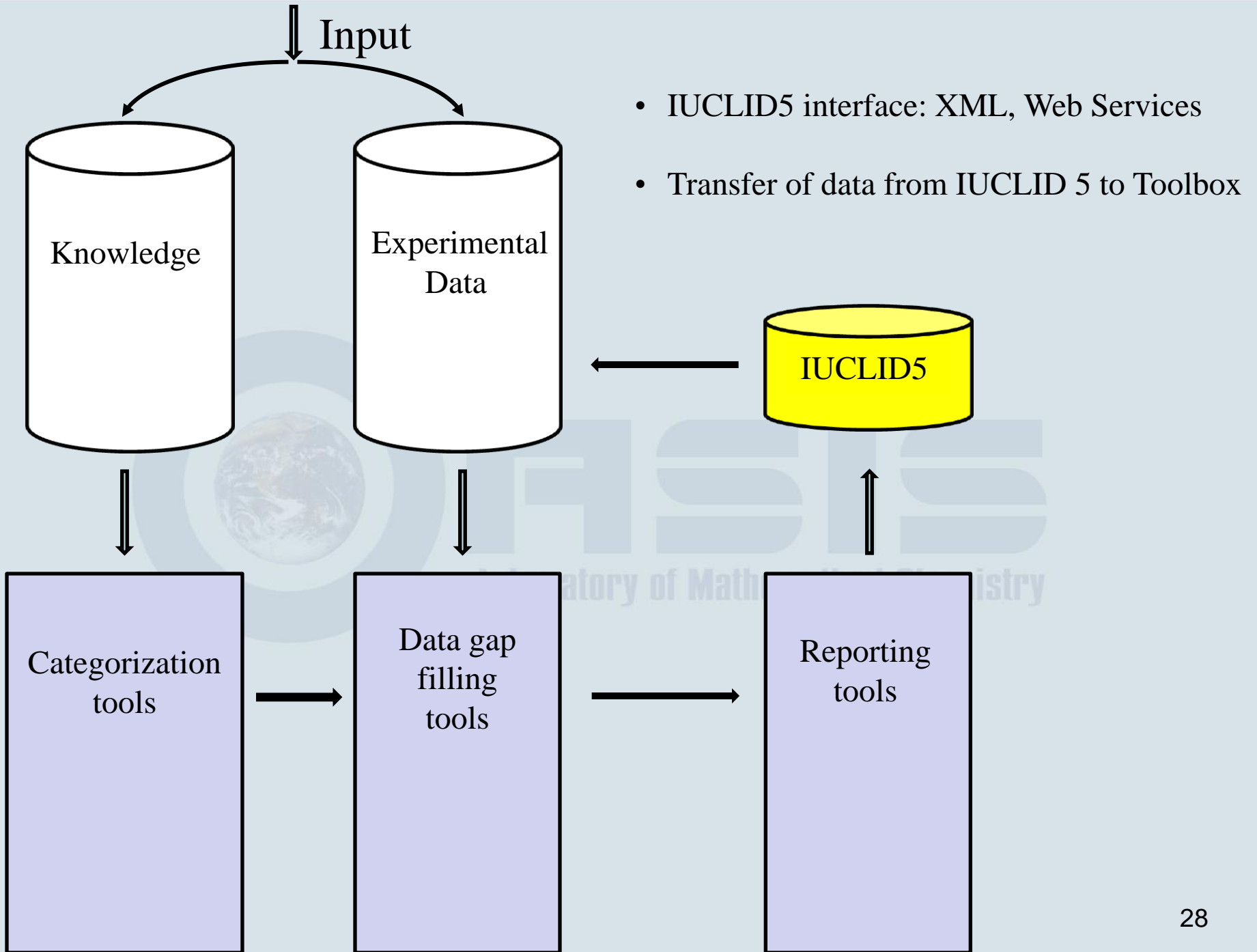


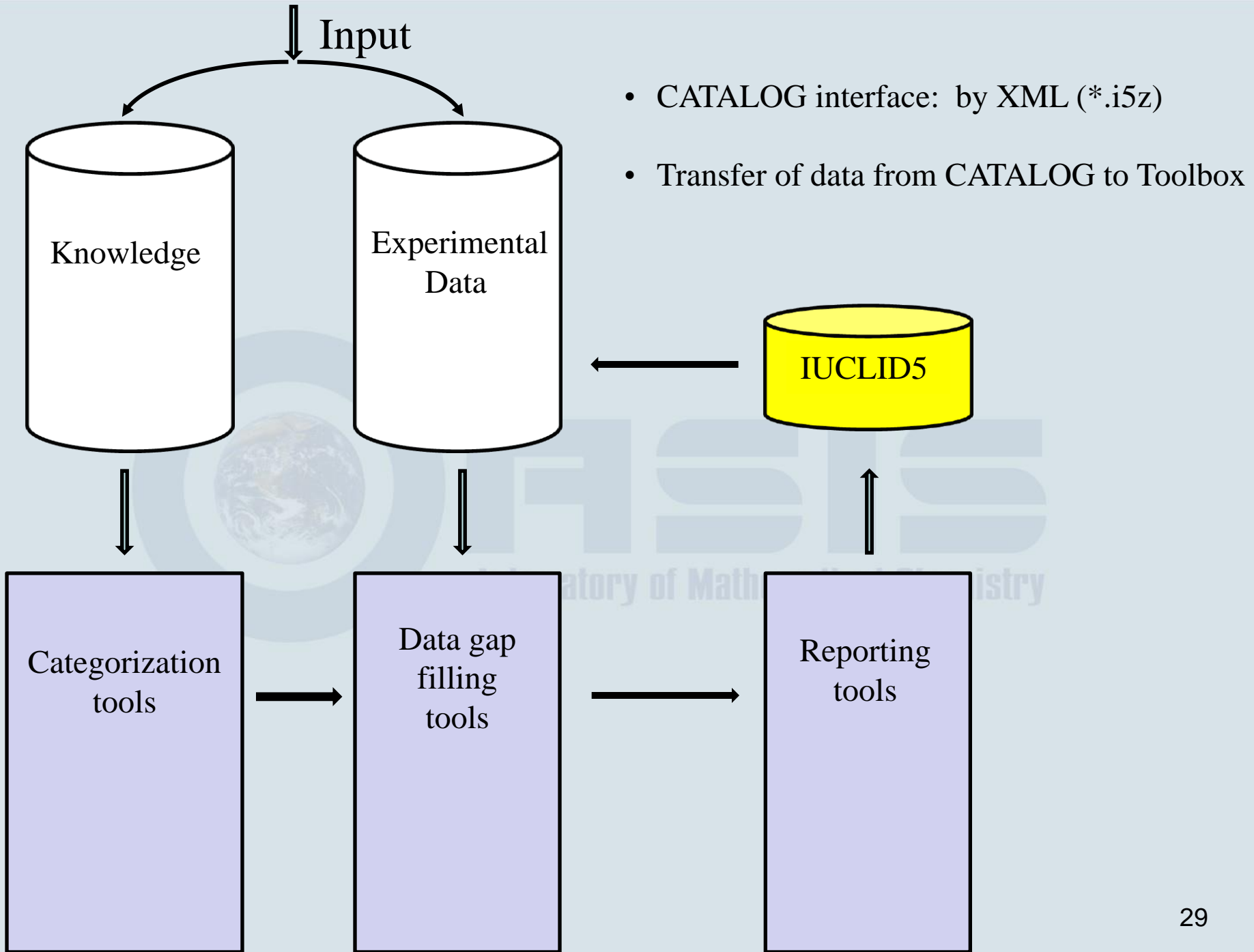
- Automated (user defined) reports:  
TPRF, TMRF (QMRF), CCRF



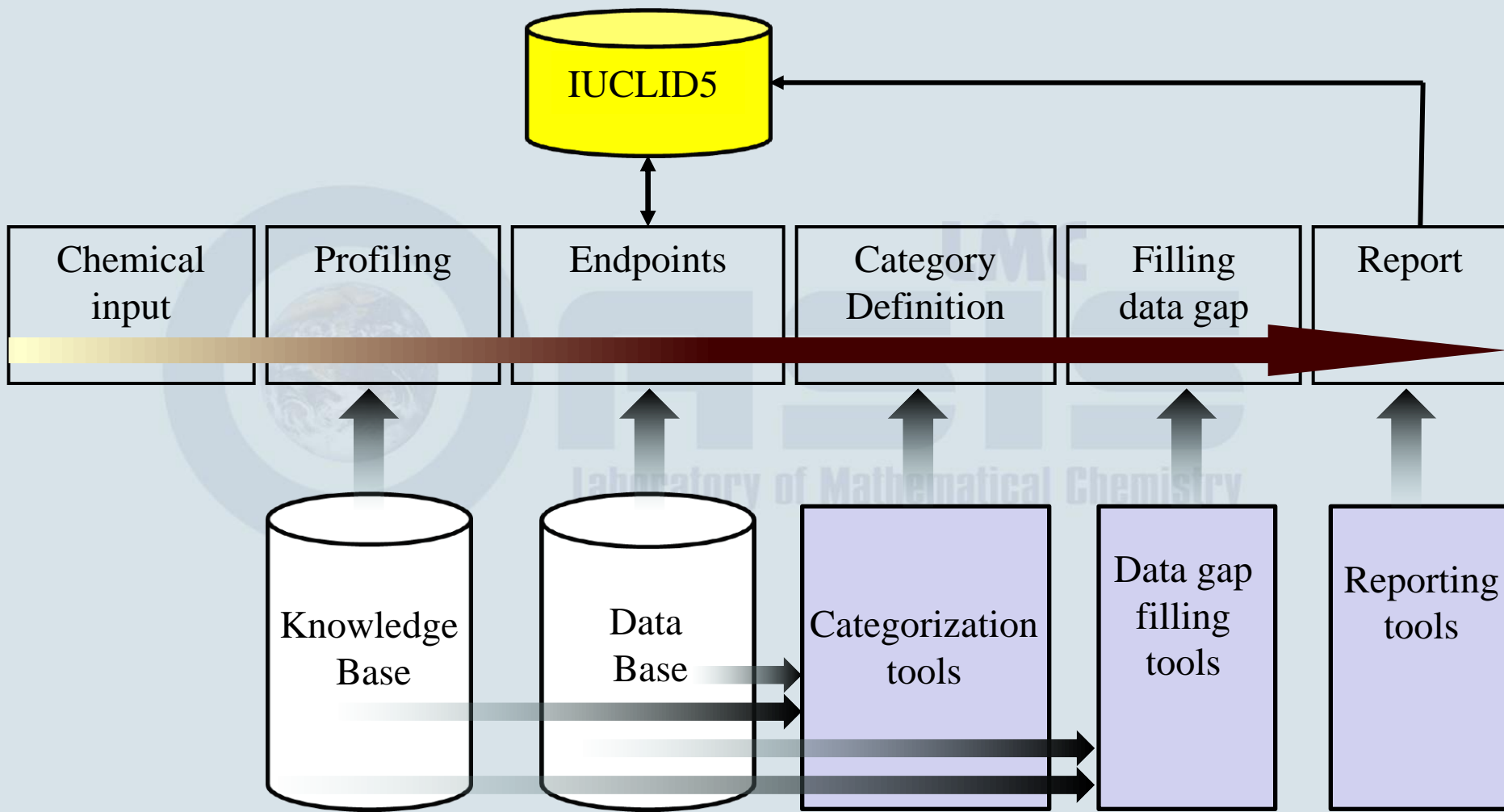


- IUCLID5 interface: XML, Web Services
- Submission of in silico predictions from Toolbox to IUCLID 5





# System Workflow



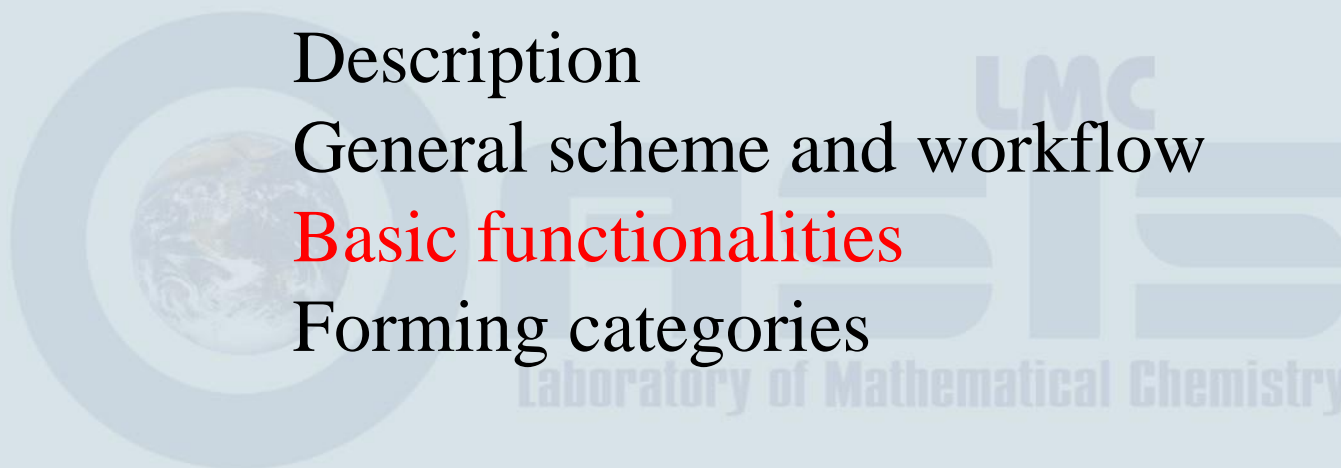
# Outlook

Description

General scheme and workflow

**Basic functionalities**

Forming categories



# Input

QSAR Toolbox 3.2.0.104 [Document\_2]

Document: Single Chemical: Chemical List

New Open Close Save CAS# Name Structure Select Delete Query ChemIDs DB Inventory List

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

Documents

- Document\_1
- Document
- Document\_2
  - CAS: 66-25-1

CCCCC=O

Structure

Filter endpoint tree... 1 [target]

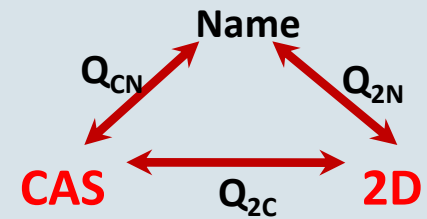
Substance Identity	
CAS Number	66-25-1
Chemical IDs	Einecs Number:20...
Chemical Name	hexanal hexaldehyde
Structural Formula	CCCCC=O
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	

... select filter type ... Create Apply

1 Document\_2

# Input

## QA of chemical structures



QSAR Toolbox 2.1.2.865 [Document\_2]

**QSAR TOOLBOX**

Input Profiling Endpoint Category Definition Data Gap Filling Report Update

Document: New, Open..., Save, Close

Single Chemical: CAS #, Chemical Name, Structure, Select from Database, Select from Inventory, Delete chemical

Chemical List: Load Database, Load Inventory, Load Chemical List

The OECD QSAR Toolbox for Grouping Chemicals into Categories

Developed by LMC, Bulgaria

Filter endpoint tree... 1

Structure

Substance Identity: 98-95-3

CAS Number: 98-95-3

Explain QA Form

Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
c1(N(=O)=O)ccccc1		1: nitrobenzene 2: benzene, nitro- 3: nitro-benzene	1: High Quality 1: Aquatic ECETOC 2: Aquatic OASIS 3: Aquatic US-EPA ECC 4: Bioaccumulation fish 5: Biodegradation OASIS 6: COSING 7: Carcinogenic Potency 8: Carcinogenicity & M 9: DSSTox 10: ECHA PR 11: EINECS 12: Genotoxicity OASIS 13: HPVC OECD 14: MITI Japan 15: Phys-chem EPTSUIT	1: High Quality 1: Aquatic ECETOC 2: Aquatic OASIS 3: Aquatic US-EPA ECC 4: Bioaccumulation fish 5: Biodegradation OASIS 6: COSING 7: Carcinogenic Potency 8: Carcinogenicity & M 9: DSSTox 10: ECHA PR 11: EINECS 12: Genotoxicity OASIS 13: HPVC OECD 14: MITI Japan 15: Phys-chem EPTSUIT	1: High Quality 1: AICS 2: Aquatic ECETOC 3: Aquatic OASIS 4: Aquatic US-EPA ECC 5: Bioaccumulation Car 6: Bioaccumulation fish 7: Biodegradation OASIS 8: COSING 9: Canada DSL 10: Carcinogenic Potency 11: Carcinogenicity & M 12: DSSTox 13: ECHA PR

# Profiling

QSAR Toolbox 3.2.0.104 [Document\_2]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Profiling Profiling Schemes

Apply New View Delete

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

Profiling methods

Select All Unselect All Invert

- Tautomers unstable
- Toxicological**
- Repeated dose (HESS)
- Custom**
- Acute Oral Toxicity
- Androgen Receptor Binding
- Cat 18. Miscellaneous cyclic chemicals
- Category 1. New Inorganics and deriv...
- Category 10: Aromatic compounds wit...
- Category 11 Aromatic compounds (nor...
- Category 12 Aromatic diamine, their d...
- Category 13 Imidazole, nitro imidazole...
- Category 14. Aromatic ring fused cycl...
- Category 15 Miscellaneous aromatic d...
- Category 16 Non-aromatic cyclic hydro...
- Category 17 Heterocyclic, cyclic compo...
- Category 18 Miscellaneous cyclic chemi...
- Category 19 Alkyl carbamodi-thioic ac...
- Category 2. Estrogen receptor (ER) an...
- Category 20 Miscellaneous non-cyclic c...
- Category 21a Vinyl amides, aldehydes
- Category 21b Vinyl amides, aldehydes
- Category 22: Alpha-substituted carbox...

Metabolism/Transformations

Select All Unselect All Invert

- Autoxidation simulator
- Autoxidation simulator (alkaline medi...
- Dissociation simulation
- Hydrolysis simulator (acidic)
- Hydrolysis simulator (basic)
- Hydrolysis simulator (neutral)
- Microbial metabolism simulator
- Rat liver S9 metabolism simulator
- Rat liver S9 metabolism simulator Last
- Skin metabolism simulator

Filter endpoint tree...

1 [target]

Structure

Inventory Affiliation

- Canada DSL
- DSSTOX
- ECHA PR
- EINECS
- METI Japan
- NICNAS
- REACH ECB
- TSCA
- US HPV Challenge...
- Not categorized
- Discrete chemical
- Aldehydes (Acute ...)

OECD HPV Chemical Categories

Substance Type

US-EPA New Chemical Categories

General Mechanistic

- Biodeg BioHC half-life (Biowin)
- Biodeg primary (Biowin 4)
- Biodeg probability (Biowin 1)
- Biodeg probability (Biowin 2)
- Biodeg probability (Biowin 5)
- Biodeg probability (Biowin 6)
- Biodeg probability (Biowin 7)
- Biodeg ultimate (Biowin 3)
- DNA binding by OASIS v.1.2
- DNA binding by OECD
- DPRA Cysteine peptide depletion

Not calculated

hours - days

Biodegrades Fast

Biodegrades Fast

Biodegrades Fast

Biodegrades Fast

Biodegrades Fast

Biodegrades Fast

days - weeks

No alert found

Schiff base formers

Schiff base former...

Schiff base former...

Low reactive

Low reactive >> Lo...

Moderate reactive

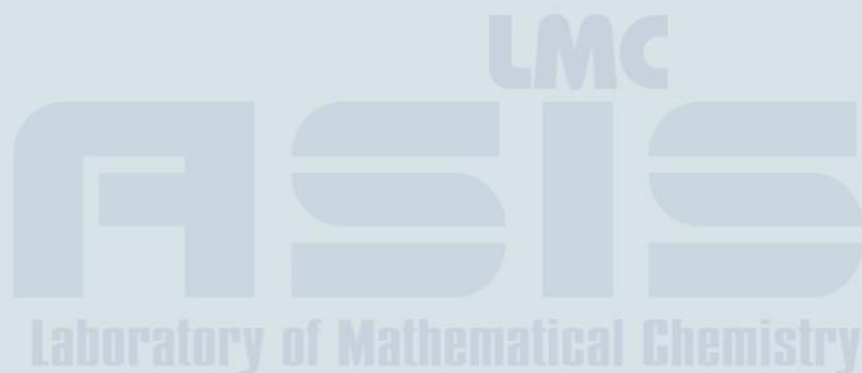
Moderate reactive ...

1 Document\_2



# Profiling

- Predefined
- General Mechanistic
- Endpoint Specific
- Empiric
- Custom



# Profiling

- **Predefined**
- General Mechanistic
- Endpoint Specific
- Empiric
- Custom

## Predefined

- Database Affiliation
- Inventory Affiliation
- OECD HPV Chemical Categories
- Substance Type
- US-EPA New Chemical Categories



LMC  
**ASIS**  
Laboratory of Mathematical Chemistry

# Profiling

- Predefined
- **General Mechanistic**
- Endpoint Specific
- Empiric
- Custom



## General Mechanistic

- Biodeg BioHC half-life (Biowin)
- Biodeg primary (Biowin 4)
- Biodeg probability (Biowin 1)
- Biodeg probability (Biowin 2)
- Biodeg probability (Biowin 5)
- Biodeg probability (Biowin 6)
- Biodeg probability (Biowin 7)
- Biodeg ultimate (Biowin 3)
- DNA binding by OASIS v.1.1
- DNA binding by OECD
- DPRA Cysteine peptide depletion
- DPRA Lysine peptide depletion
- Estrogen Receptor Binding
- Hydrolysis half-life (Ka, pH 7)(Hydrowin)
- Hydrolysis half-life (Ka, pH 8)(Hydrowin)
- Hydrolysis half-life (Kb, pH 7)(Hydrowin)
- Hydrolysis half-life (Kb, pH 8)(Hydrowin)
- Hydrolysis half-life (pH 6.5-7.4)
- Ionization at pH = 1
- Ionization at pH = 4
- Ionization at pH = 7.4
- Ionization at pH = 9
- Protein binding by OASIS v1.1
- Protein binding by OECD
- Protein binding potency
- Superfragments
- Toxic hazard classification by Cramer (original)
- Toxic hazard classification by Cramer (with extensions)
- Ultimate biodeg

# Profiling

- Predefined
- General Mechanistic
- **Endpoint Specific**
- Empiric
- Custom



## Endpoint Specific

- ✓ Acute aquatic toxicity classification by Verhaar
- ✓ 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
- ✓ DNA alerts for AMES, MN and CA 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 skin sensitization by OASIS v1.1
- ✓ rTER Expert System ver.1 - USEPA
- ✓ Skin irritation/corrosion Exclusion rules by BfR
- ✓ Skin irritation/corrosion Inclusion rules by BfR

# Profiling

- Predefined
- General Mechanistic
- Endpoint Specific
- **Empiric**
- **Custom**



**ALISIS**  
Laboratory of Mathematical Chemistry

## Empiric

- Chemical elements
- Groups of elements
- Lipinski Rule Oasis
- Organic functional groups
- Organic functional groups (nested)
- Organic functional groups (US EPA)
- Organic functional groups, Norbert Haider (checkmol)
- Tautomers unstable

## Toxicological

- Repeated dose (HESS)

## Custom

- aldehydes
- Anilines

# Profiling

- Predefined
- General Mechanistic
- Endpoint Specific
- Empiric
- Custom

- **Molecular transformations**

## Metabolism/Transformations

### Documented

- Observed Mammalian metabolism
- Observed Microbial metabolism
- Observed Rat In vivo metabolism
- Observed Rat Liver S9 metabolism

### Simulated

- Autoxidation simulator
- Autoxidation simulator (alkaline medium)
- Dissociation simulation
- Hydrolysis simulator (acidic)
- Hydrolysis simulator (basic)
- Hydrolysis simulator (neutral)
- Microbial metabolism simulator
- Rat liver S9 metabolism simulator
- Skin metabolism simulator

# Endpoint

QSAR Toolbox 3.2.0.104 [Document\_2]

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

Data Import Export Delete Tautomerize

Gather Import IUCLID5 Export IUCLID5 Database Inventory Database

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

Databases

Select All Unselect All Invert About

- Phys-chem EPISUITE
- Environmental Fate and Transport**
  - Bioaccumulation Canada
  - Bioaccumulation fish CEFIC LRI
  - Bioconcentration NITE
  - Biodegradation in soil OASIS
  - Biodegradation NITE
  - Biota-Sediment Accumulation Factor U
  - ECHA CHEM
  - ECOTOX
  - Hydrolysis rate constant OASIS
  - KM database Environment Canada
  - Phys-chem EPISUITE
- Ecotoxicological Information**
  - Aquatic ECETOC
  - Aquatic Japan MoE
  - Aquatic OASIS
  - ECHA CHEM
  - ECOTOX
- Human Health Hazards**
  - Acute oral toxicity\_exp data
  - Bacterial mutagenicity ISSSTY
  - Carcinogenic Potency Database (CPDB)
  - Carcinogenicity&mutagenicity ISSCAN

Inventories

Select All Unselect All Invert About

- Canada DSL
- COSING
- DSSTOX
- ECHA PR
- EINECS
- HPVC OECD
- METI Japan
- NICNAS
- REACH ECB
- TSCA
- US HPV Challenge Program

Filter endpoint tree... 1 [target]

Structure

Substance Identity

Physical Chemical Properties (1/6) M: 131 °C, 4.41, 1....

Environmental Fate and Transport (1/4) M: 50 %, 21.6 Pa...

Ecotoxicological Information

- Aquatic Toxicity
  - Growth (1/1) M: 152 mg/L
  - Immobilisation
  - Intoxication (1/12) M: 16 mg/L, 18 mg...
  - Mortality
    - EC50 (1/25) M: 17.8 mg/L, 9.79...
    - LD50
    - MRC50
    - NR-LETH (1/1) M: 16 mg/L
- Physiology
- Undefined Effect
- Sediment Toxicity
- Terrestrial Toxicity (1/4) M: 38 milliliters per...

Human Health Hazards (1/1) M: 4.89E3 mg/kg

- Acute Toxicity (1/1) M: 4.89E3 mg/kg
- Carcinogenicity
- Developmental Toxicity / Teratogenicity
- Genetic Toxicity
- Immunotoxicity
- Irritation / Corrosion (1/1) M: Irritant

# Endpoint

- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazard



LMC  
**ASIS**  
Laboratory of Mathematical Chemistry



# Endpoint

- **Physical Chemical Properties**
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazard

- ▣ **Physical Chemical Properties**
  - ▣ Chemical Reactivity COLIPA
  - ▣ ECHA CHEM
  - ▣ Experimental pKa
  - ▣ GSH Experimental RC50
  - ▣ Phys-chem EPISUITE



LMC  
**ASIS**  
Laboratory of Mathematical Chemistry

# Endpoint

- Physical Chemical Properties
- **Environmental Fate and Transport**
- Ecotoxicological Information
- Human Health Hazard



**ALICE**  
Laboratory of Mathematical Chemistry

- ▣ **Environmental Fate and Transport**
  - ▣ Aquatic US-EPA ECOTOX
  - ▣ Bioaccumulation Canada
  - ▣ Bioaccumulation fish CEFIC LRI
  - ▣ Bioconcentration NITE
  - ▣ Biodegradation in soil OASIS
  - ▣ Biodegradation NITE
  - ▣ Biota-Sediment Accumulation Factor US-EPA
  - ▣ ECHA CHEM
  - ▣ Hydrolysis rate constant OASIS
  - ▣ kM database Environment Canada
  - ▣ Phys-chem EPISUITE
  - ▣ Terrestrial US-EPA ECOTOX

# Endpoint

- Physical Chemical Properties
- Environmental Fate and Transport
- **Ecotoxicological Information**
- Human Health Hazard

- ▣ Ecotoxicological Information
  - ▣ Aquatic ECETOC
  - ▣ Aquatic Japan MoE
  - ▣ Aquatic OASIS
  - ▣ Aquatic US-EPA ECOTOX
  - ▣ ECHA CHEM
  - ▣ Terrestrial US-EPA ECOTOX



LMC  
**OASIS**  
Laboratory of Mathematical Chemistry

# Endpoint

- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- **Human Health Hazard**



Labo

- ▣ **Human Health Hazards**
  - ▣ Bacterial mutagenicity ISSSTY
  - ▣ Carcinogenic Potency Database (CPDB)
  - ▣ Carcinogenicity&mutagenicity ISSCAN
  - ▣ Cell Transformation Assay ISSCTA
  - ▣ Dendritic cells COLIPA
  - ▣ Developmental toxicity ILSI
  - ▣ ECHA CHEM
  - ▣ Estrogen Receptor Binding Affinity OASIS
  - ▣ Eye Irritation ECETOC
  - ▣ Genotoxicity OASIS
  - ▣ Keratinocyte gene expression Givaudan
  - ▣ Micronucleus ISSMIC
  - ▣ Micronucleus Oasis
  - ▣ MUNRO non-cancer EFSA
  - ▣ Rep Dose Tox Fraunhofer ITEM
  - ▣ Repeated Dose Toxicity HESS
  - ▣ Rodent Inhalation Toxicity Database
  - ▣ Skin irritation
  - ▣ Skin sensitization
  - ▣ Skin sensitization ECETOC
  - ▣ Terrestrial US-EPA ECOTOX
  - ▣ Toolbox Database example
  - ▣ Toxicity Japan MHLW
  - ▣ ToxRefDB US-EPA
  - ▣ Yeast estrogen assay database University of Tennessee-Knoxville (USA)

# Endpoint

QSAR Toolbox 3.1.0.

QSAR TO

Data

Gather

Import

Select All

Unse

Ecotoxicologica

Aquatic ECETO

Aquatic Japan

Aquatic OASIS

Aquatic US-EPA ECOTOX

ECHA CHEM

Terrestrial US-EPA ECO

Toolbox Database exar

Verical import\_ChemID

Human Health Hazards

Bacterial mutagenicity ISSSTY

Carcinogenic Potency Database (CPDB)

Carcinogenicity&mutagenicity ISSCAN

Cell Transformation Assay ISSCTA

Dendritic cells COLIPA

Developmental toxicity ILSI

ECHA CHEM

**About**

**Name**  
Aquatic US-EPA ECOTOX

**Short description**  
ECOTOX is a comprehensive database, which provides information on adverse effects of single chemical stressors to ecologically relevant aquatic and terrestrial species. ECOTOX includes more than 765,032 test records covering 9,731 aquatic and terrestrial species and 9,562 chemicals. The primary source of ECOTOX data is the peer-reviewed literature, with test results identified through comprehensive searches of the open literature. The aquatic data were originally presented in a separate EPA database called AQUIRE (AQUatic Information Retrieval). AQUIRE was established in 1981 by the EPA and was maintained by the Mid-Caribbean Ecology Division of the National Health and

**Disclaimer**  
You should consult the original scientific paper to ensure an understanding of the context of the data retrieved from the ECOTOX database. ECOTOX attempts to be comprehensive, our

**Donator(s)**  
United States Environmental Protection Agency (EPA), USA

**Author(s)**

**Website**  
<<http://cfpub.epa.gov/ecotox/>>

**Details**

Number of chemicals	7,894
Number of data	337,204
Number of endpoints	147
Name of endpoints	BCF,NOEC,EC50,LC50,LOEC,IC50,LD50,EC2
Version	

Close

Select all

Unselect all

Invert selection

About

Substances Identity

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Aquatic Toxicity

Accumulation (1/9)

Avoidance (1/3)

Behavior (2/10)

Biochemistry (2/24)

Cell(s)

Development (1/6)

Enzyme(s) (1/4)

Feeding Behavior (1/1)

## Task C - ECHA CHEM database

C. Expansion of the repertoire of databases including data content normalization

### Development of methodology to explore ECHA CHEM database

#### QSAR Toolbox data and ECHA CHEM contribution

Endpoint tree	QSAR Toolbox (without ECHA CHEM)		ECHA CHEM contribution	
	Substances	Data	Substances	Data
Phys-chem properties	38 000	71 000	5 700 (15%)	85 000 (120%)
Environmental	4 500	58 000	4 000 (88%)	28 000 (48%)
Ecotoxicological	8 500	510 000	4 600 (54%)	95 000 (18%)
Human health	15 000	550 000	5 400 (36%)	140 000 (25%)
<b>Total</b>	<b>55 000</b>	<b>1 200 000</b>	<b>6 200 (9%)</b>	<b>350 000 (29%)</b>

# QSAR Toolbox databases

## *Contribution of REACH dissemination data ECHA CHEM database*

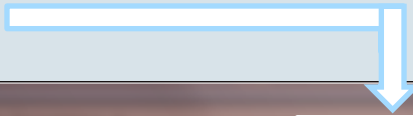
Endpoint tree	ECHA CHEM		Total	
	Substances	Data	Substances	Data
Substance Identity	6188	301 868	92 854	1 538 014
Physical Chemical Properties	5 738	86 453	41931	156 528
Environmental Fate and Transport	4 001	25 759	7581	85 585
Ecotoxicological Information	4 626	94 009	11 895	608 065
Human Health Hazards	5 426	95 647	18 813	687 836

# Inventories

Inventories	Number of chemicals
AICS	38760
COSING	1314
DSSTOX	8606
ECHA PR	142619
EINECS	72561
HPVC EU	4843
METI Japan	16133
REACH ECB	74074
US HPV Challenge Program	9125
TSCA 2005	65709
Canada DSL	22017
<b>Total number</b>	<b>197015</b>



# Category definition



QSAR Toolbox 3.2.0.104 [Document\_2]

**QSAR TOOLBOX**

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

Categorize Delete

Define Subcategorize Combine Clustering Delete Delete All

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

Grouping methods

- Predefined
  - Database Affiliation
  - Inventory Affiliation
  - OECD HPV Chemical Categories
  - Substance Type
  - US-EPA New Chemical Categories**
- General Mechanistic
  - Biodeg BioHC half-life (Biowin)
  - Biodeg primary (Biowin 4)
  - Biodeg probability (Biowin 1)
  - Biodeg probability (Biowin 2)
  - Biodeg probability (Biowin 5)
  - Biodeg probability (Biowin 6)
  - Biodeg probability (Biowin 7)
  - Biodeg ultimate (Biowin 3)
  - DNA binding by OASIS v. 1.2
  - DNA binding by OECD
  - DPRAs Cysteine peptide depletion
  - DPRAs Lysine peptide depletion
  - Estrogen Receptor Binding
  - Hydrolysis half-life (Ka, pH 7)(Hydrowin)
  - Hydrolysis half-life (Ka, pH 8)(Hydrowin)
  - Hydrolysis half-life (Kb, pH 7)(Hydrowin)
  - Hydrolysis half-life (Kb, pH 8)(Hydrowin)
  - Hydrolysis half-life (pH 6.5-7.4)
  - Ionization at pH = 1
  - Ionization at pH = 4
  - Ionization at pH = 7.4

Defined Categories

- Document\_2
  - [748] Aldehydes (Acute toxicity) (US-EPA New Chemical Categories)**

Filter endpoint tree... 1 [target] 2 3 4 5 6

Structure	1 [target]	2	3	4	5	6
<ul style="list-style-type: none"> <li>Substance Identity</li> <li>Physical Chemical Properties (456/1019) M: 131 °C, 4.41, 1...</li> <li>Environmental Fate and Transport                             <ul style="list-style-type: none"> <li>Bioaccumulation: Aquatic (2/15)</li> <li>Bioaccumulation: Terrestrial</li> <li>Biodegradation (34/34) M: 50 % M: 71 % M: 2 %</li> <li>Photodegradation (44/77) M: 1.57E-14 cm3/... M: 5.8E-16 cm3/m...</li> <li>Stability in Water</li> <li>Transport and Distribution Between Environm... (36/36) M: 21.6 Pa-m3/mole M: 0.0341 Pa-m3/... M: 0.568 Pa-m3/mole M: 14.9 Pa-m3/mole</li> </ul> </li> <li>Ecotoxicological Information                             <ul style="list-style-type: none"> <li>Aquatic Toxicity (157/1799) M: 17.8 mg/L, 9.79... M: 23.8 mg/L, 27.4... M: 0.13 mg/L, 0.22... M: 2.5 mg/L, 3 mg... M: 0.0</li> <li>Sediment Toxicity</li> <li>Terrestrial Toxicity (42/754) M: 38 milliliters per... M: 0.05 Active Ingr... M: 17.3(12.1;19.7) ... M: 353 mg/L</li> </ul> </li> <li>Human Health Hazards                             <ul style="list-style-type: none"> <li>Acute Toxicity (10/31) M: 1 mg/L, ≈58(39;... M: Positive, Positiv...</li> <li>Carcinogenicity (29/124) M: 40 mg/L, 20 mg...</li> <li>Developmental Toxicity / Teratogenicity (3/13)</li> <li>Genetic Toxicity (242/1621) M: Positive, Negati...</li> <li>Immunotoxicity</li> <li>Irritation / Corrosion (29/37) M: 2.54</li> <li>Neurotoxicity</li> <li>Repeated Dose Toxicity (39/5306) M: 82 mg/kg/day, ... M: 10 mg/kg/day, ...</li> </ul> </li> </ul>						

748 Aldehydes (Acute toxicity) (US-EPA New Chemical Categories) 1 3/0/0

# Data Gap Filling

**QSAR TOOLBOX 3.2.0.104 [Document\_2]**

**Navigation:** Input | Profiling | Endpoint | Category Definition | **Data Gap Filling** | Report

**Left Sidebar (Filling):**

- Data Gap Filling Method:**
  - Read-across
  - Trend analysis
  - (Q)SAR models
- Target Endpoint:**
  - Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata(Vertebrates) Actinopterygii(Fish) Pimephales promelas

**Central Plot:**

making a linear approach for the chemicals, Observed target value: 17.8 mg/L, Predicted target value: 7.24 mg/L

Y-axis: LC50 (obs.), log(f mol/L)

X-axis: log Kow

Descriptor X: log Kow

**Context Menus:**

- Selection navigation:** Go back, Go forward, Go to first, Go to last
- Gap filling approach:** Read-across, Trend analysis
- Descriptors/data:** Make active descriptor, Remove active descriptor, Collect data, Change descriptor units ..., Edit descriptor options ...
- Model/(Q)SAR:** Show domain, Save model, Save domain as category, Save JRC XML QMRF, Calculate Q2

**Right Sidebar (Select/filter data):**

- Select/filter data
- Subcategorize
- Mark chemicals by WS
- Mark chemicals by descriptor value
- Mark outlier points
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked chemicals/points
- Clear existing marks

**Bottom Bar:** 748 Aldehydes (Acute toxicity) (US-EPA New Chemical Categories) 1 | Create predic | 0/100 | 3/1/0

# Data Gap Filling

QSAR Toolbox 2.1.2.865 [Document]

**QSAR TOOLBOX**

Input

Apply

Control

Data Gap Filling Method

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

Target Endpoint

Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata(Vertebrates) Actinopterygii(Fish) Pimephales promelas

Set usage of data per chemical:

- All
- Minimal
- Maximal
- Average**
- Median(s)
- Lower median
- Higher median
- Mode(s)
- Lowest mode
- Highest mode

Calculation options

- Data usage
- Prediction approach options
- Use target data for prediction
- Set level of significance

Visual options

- Set units in figure title
- Set axes ranges
- Show confidence range
- Show intercorrelations

Information

- Details of focused chemical
- Details of focused point
- Details of target chemical
- Differences to target
- All points within region
- Show legend

Miscellaneous

- Print chemicals
- Save chemicals to SMI
- Copy picture

Accept prediction

Return to matrix

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

Structure

OK Cancel

Descriptors Prediction Adequacy

Trend analysis prediction of LC50, making a linear approximation of observed target values

LC50 (obs.), log(1/mol/L)

log Kow

70 analogue chemicals, value: 4.14 log(1/mol/L), log Kow

3(2.2-2.5)... M: 150 mg/L M: 0.014 mg/L, ... M: 2

678 Aldehydes (Acute toxicity) (US-EPA New Chemical Categories Data gap filling)

# Report

The screenshot displays the QSAR Toolbox 3.2.0.104 software interface. The main window is titled "Prediction [1]". The interface includes a top menu bar with options: Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below the menu bar is a toolbar with icons for Create, Print, Close, Save as, Register, Unregister, Update, Clone, and Design. The left sidebar shows a tree view with "Available data to report" containing "Predictions", "(Q)SARs", and "Categories". Below this is "Available report templates" with "Standard (predefined)" and "Custom (user defined)" sections. The main content area displays the report title "Prediction of LC50 for hexanal" and the page number "1 / 28". The report content includes the heading "QSAR Toolbox prediction for single chemical" and a paragraph of text: "The template of the current report is based on 'GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS MODELS' published by OECD (September, 2007) and 'GUIDANCE ON INFORMATION REQUIREMENTS AND CHEMICAL SAFETY ASSESSMENT / CHAPTER R.6: QSARS AND GROUPING OF CHEMICALS' published by ECHA (May, 2008). The report provides information about the target substance, chemical characteristics used for the grouping, the resulting boundaries of the group of chemicals (applicability". The status bar at the bottom shows "748 Aldehydes (Acute toxicity) (US-EPA New Chemical Categories) 1" and "3/0/0".

QSAR Toolbox 3.2.0.104 [Document\_2]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Reports Repository

Create Print Close Save as Register Unregister Update Clone Design

Available data to report

- Predictions
- (Q)SARs
- Categories

Available report templates

- Standard (predefined)
  - QSAR Toolbox Prediction Report (TPRF v.3.2)
- Custom (user defined)
  - Editable copy of QSAR Toolbox Prediction Rep

Prediction [1]

Prediction of LC50 for hexanal 1 / 28

### QSAR Toolbox prediction for single chemical

*The template of the current report is based on "GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS MODELS" published by OECD (September, 2007) and "GUIDANCE ON INFORMATION REQUIREMENTS AND CHEMICAL SAFETY ASSESSMENT / CHAPTER R.6: QSARS AND GROUPING OF CHEMICALS" published by ECHA (May, 2008). The report provides information about the target substance, chemical characteristics used for the grouping, the resulting boundaries of the group of chemicals (applicability*

748 Aldehydes (Acute toxicity) (US-EPA New Chemical Categories) 1 3/0/0

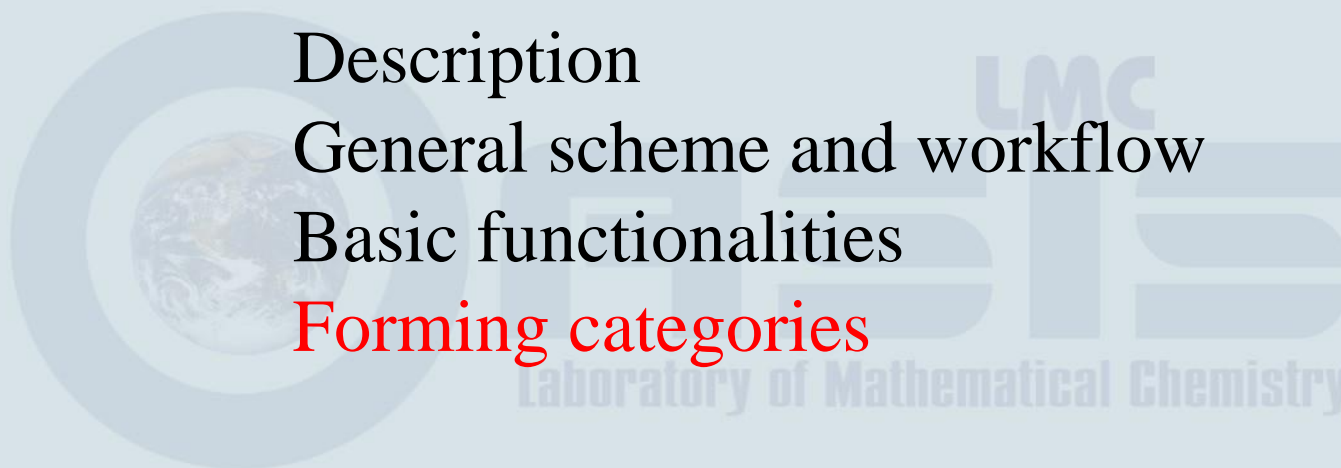
# Outlook

Description

General scheme and workflow

Basic functionalities

Forming categories



# How to build categories?



# Basic guidance for category formation

## Recommended categorization phases:

1. Phase I. Endpoint non-specific - structure-related profilers (primary categorization)
2. Phase II. Endpoint specific profilers (for subcategorization) – based on endpoint driving interaction mechanisms
3. Phase II. Additional structure-related profilers, to further eliminate dissimilar chemicals (to increase the consistency of category)

Laboratory of Mathematical Chemistry

## Recommended Categorization Phases

### Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping  
Endpoint Non-specific

Repeating Phase I due to Multifunctionality of chemicals





## Recommended Categorization Phases

### Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping  
Endpoint Non-specific

Repeating Phase I due to Multifunctionality of chemicals

### Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Mode of action –acute aquatic toxicity
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

Subcategorization  
Endpoint Specific

Metabolism accounted for

## Recommended Categorization Phases

### Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Broad grouping  
Endpoint Non-specific

Repeating Phase I due to Multifunctionality of chemicals

### Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Mode of action –acute aquatic toxicity
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

Subcategorization  
Endpoint Specific

Metabolism accounted for

### Phase III. Eliminating dissimilar chemicals

Apply Phase I categorization

Subcategorization  
Endpoint Specific

# Basic guidance for category formation

## Suitable categorization phases:

1. Phase I. Endpoint non-specific - structure-related profilers (primary categorization)
2. Phase II. Endpoint specific profilers (for subcategorization) – based on endpoint driving interaction mechanisms
3. Phase II. Additional structure-related profilers, to further eliminate dissimilar chemicals (to increase the consistency of category)

## Performing categorization:

1. The categorization phases should be applied successively
2. The application order of the phases depend on the specificity of the data gap filling performed (data availability, endpoint specificity)
3. More categories of same phase could be used in forming categories
4. Some of the phases could be skipped if consistency of category members is reached
5. Subcategorization should be applied at Data gap filling stage

# Example – Ecotoxicological Information

## Grouping methods

### General Mechanistic

- DNA binding by OASIS
- DNA binding by OECD
- Estrogen Receptor Binding
- Protein binding by OASIS
- Protein binding by OECD
- Protein Binding Potency
- Superfragments
- Toxic hazard classification by Cramer (original)
- Toxic hazard classification by Cramer (with extensions)

### Endpoint Specific

- Acute aquatic toxicity classification by Verhaar
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Bioaccumulation – metabolism alerts
- Bioaccumulation – metabolism half-lives
- Biodegradation fragments (BioWIN MITI)
- Eye irritation/corrosion Exclusion rules by BfR
- Eye irritation/corrosion Inclusion rules by BfR
- Micronucleus alerts by Benigni/Bossa
- Mutagenicity/Carcinogenicity alerts by Benigni/Bossa
- Oncologic Primary Classification
- Skin irritation/corrosion Exclusion rules by BfR
- Skin irritation/corrosion Inclusion rules by BfR

### Empiric

- Chemical elements
- Groups of elements
- Lipinski Rule Oasis
- Organic functional groups
- Organic functional groups (nested)
- Organic functional groups (US EPA)
- Organic functional groups, Norbert Haider (checkmol)
- Structure similarity

### Custom

- Aldehydes

## Acute Aquatic Toxicity

Mode of Action

Structural similarity (not for narcotics)

Data Gap Filling Approach

Trend analysis, External QSARs

# Example – Human Health Hazard

## Grouping methods

### General Mechanistic

- DNA binding by OASIS
- DNA binding by OECD
- Estrogen Receptor Binding
- Protein binding by OASIS
- Protein binding by OECD
- Protein Binding Potency
- Superfragments
- Toxic hazard classification by Cramer (original)
- Toxic hazard classification by Cramer (with extensions)

### Endpoint Specific

- Acute aquatic toxicity classification by Verhaar
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Bioaccumulation – metabolism alerts
- Bioaccumulation – metabolism half-lives
- Biodegradation fragments (BioWIN MITI)
- Eye irritation/corrosion Exclusion rules by BfR
- Eye irritation/corrosion Inclusion rules by BfR
- Micronucleus alerts by Benigni/Bossa
- Mutagenicity/Carcinogenicity alerts by Benigni/Bossa
- Oncologic Primary Classification
- Skin irritation/corrosion Exclusion rules by BfR
- Skin irritation/corrosion Inclusion rules by BfR

### Empiric

- Chemical elements
- Groups of elements
- Lipinski Rule Oasis
- Organic functional groups
- Organic functional groups (nested)
- Organic functional groups (US EPA)
- Organic functional groups, Norbert Haider (checkmol)
- Structure similarity

### Custom

- Aldehydes

## Mutagenicity/Carcinogenicity

Mechanism

Structural similarity

Data Gap Filling Approach

Trend analysis, External SARs