

**ambit**



# **AMBIT – open source tools to integrate and explore chemical substance databases**

Dr. Nina Jeliaskova

Ideaconsult Ltd. , Sofia, BG

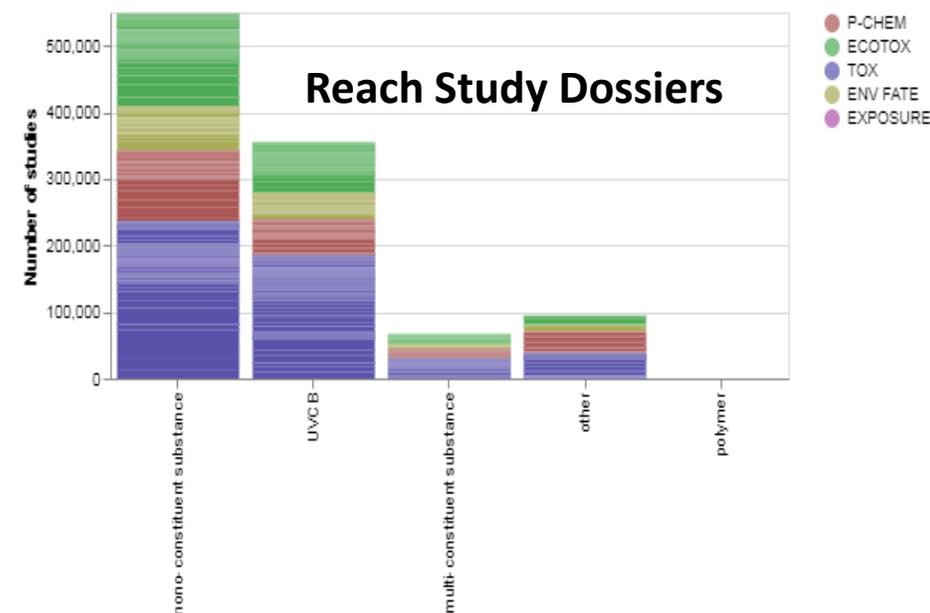


[www.ideaconsult.net](http://www.ideaconsult.net)

# Background: AMBIT – open source tools to integrate chemical substance databases



- Developed within a CEFIC Long-Range Initiative (LRI) since 2005
- <http://cefic-lri.org/toolbox/ambit/> ( links, video & guides)
- Continuously developed and extended through various projects
  - (EU FP7 & H2020 , industry);
- Chemical structure database with support for **substances**
  - including nanomaterials, advanced materials
- Data analysis & prediction tools;
- Tools to process and import data (doi: 10.3390/nano10101908); Export in various formats ;
- Supports **REST API** since 2010 (doi:10.1186/1758-2946-3-18)
- **Read Across Workflow** <https://ambitlri.ideaconsult.net> with REACH dossiers (and more)
- **Open source** <https://ambit.sf.net>



>1400 registered users  
> 900 organisations

# What's in a substance ?



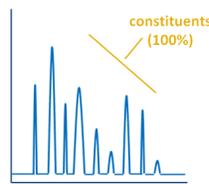
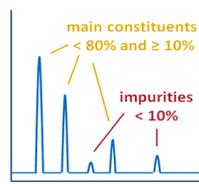
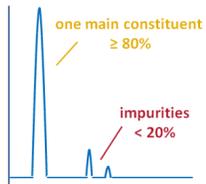
## IUPAC definition:

a substance is a matter of constant composition, best characterised by the entities (molecules, formula units, atoms) it is composed of.



## ECHA Substance Identification

A substance is a chemical element and its compounds in the natural state or the result of a manufacturing process.



Substance Name	Substance UUID	Substance Type	Public name	Reference substance UUID	Owner	Info
50-00-0_Formaldehyde_TRA0001_CLN test substance_TST_DE00_REACH P3	IUCS-410e026d-f...	UVCB	Formaldehyde	ECBS-053aa8c4-...	Clariant Produkte (Deutschland) GmbH / Sulzbach am Taunus / Germany	
<b>Composition name:</b> Benzoic Acid,Acidum Benzoicum (5 % (w/w)) <b>Composition UUID:</b> L-4711194d-336f-497e-9667-34cb5852627d <b>Purity of IUC Substance:</b> 75-90 % (w/w)						
Type	Name	EC No.	CAE No.	Typical concentration	Concentration ranges	Structure
Additive	1-Ethoxy-2-(2-Ethoxyethoxy)Ethane, 1,1'-Oxybis(2-Ethoxyethane),Diethylene Glycol Diethyl Ether,Bis(2-Ethoxyethyl) Ether,Diethylene_glycol_diethyl_ether,Ethane, 1,1'-Oxybis(2-Ethoxy-	203-903-7	112-36-7	ca. 0.5 % (w/w)	$>= 0.2\%$ (w/w) $<= 0.6\%$ (w/w)	Also contained in... 
Constituent	Benzoic Acid,Acidum Benzoicum	200-618-2	66-85-0	5 % (w/w)	0 % (w/w) 0 % (w/w)	Also contained in... 
Constituent	Salicylaldehyde,2-Hydroxybenzaldehyde,Benzaldehyde, 2-Hydroxy-	201-961-0	90-02-8	ca. 5 % (w/w)	$>= 2\%$ (w/w) $<= 8\%$ (w/w)	Also contained in... 
Constituent	Methyl 4-Methylbenzoate,Methyl P-Toluate,Benzoic Acid, 4-Methyl-, Methyl Ester	202-784-1	96-75-2	ca. 10 % (w/w)	$>= 8\%$ (w/w) $<= 15\%$ (w/w)	Also contained in... 
Constituent	Formaldehyde	200-001-8	50-00-0	ca. 80 % (w/w)	$>= 75\%$ (w/w) $<= 90\%$ (w/w)	Also contained in... 
Impurity	2-Hydroxy-1-Naphthaldehyde,1-Naphthaldehyde, 2-Hydroxy-, 1-Naphthalenecarboxaldehyde, 2-Hydroxy-	211-002-0	708-08-5	ca. 1 % (w/w)	$>= 0.5\%$ (w/w) $<= 2\%$ (w/w)	Also contained in... 

Additive

Constituent

Main constituent

Impurity

# REACH substance definition as implemented in IUCLID

## Example: mono-constituent

- 1.2 Composition
  - Monoethylene glycol dimethyl ether
  - Crude Monoglyme 1
  - Crude Monoglyme 2

Three different compositions

EU: REACH

>= 97.1 % (w/w)

**Constituents**

1,2-dimethoxyethane / 1,2-dimethoxyethane / 110-71-4 / 203-794-9, ca. 97.8 % (w/w), >= 97.1 - <= 97.9 % (w/w)
---

**Impurities**

1,4-dioxane / 1,4-dioxane / 123-91-1 / 204-661-8, ca. 1.4 % (w/w), >= 1.0 - <= 2.0 % (w/w)
2-methyl-1,3-dioxolane / 2-methyl-1,3-dioxolane / 497-26-7 / 207-841-4, ca. 0.8 % (w/w), >= 0.5 - <= 1.5 % (w/w)
unknown organic impurities / unknown, < 0.2 % (w/w), <= 0.5 % (w/w)

**Additives**

The REACH definition of a substance encompasses all forms of substances and materials on the market, including nanomaterials; and may have complex composition.

**Mono-constituent:** A substance with one main constituent.

**Multi-constituent:** A substance with two or more main constituents.

**Main constituent:** A constituent, not being an additive or impurity, in a substance that makes up a significant part of that substance. Contributes to the naming of the substance. Concentration of the main constituent(s) = purity of the substance.

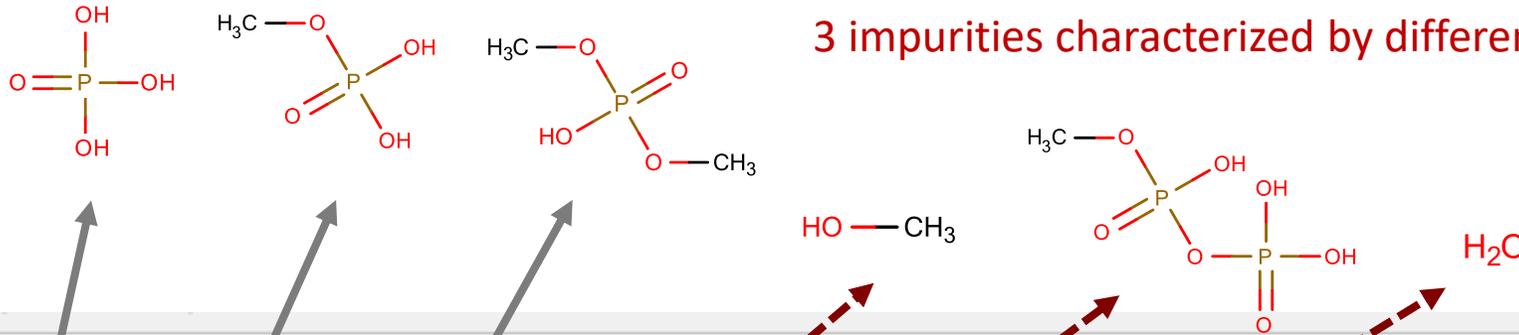
**Additive:** A substance that has been intentionally added to stabilise the substance. Contributes to the substance composition.

**Impurity:** An unintended constituent present in a substance, as produced. Does not contribute to the naming of the substance

# REACH substance definition implemented in IUCLID

## Example: multi-constituent

The substance has 3 constituents and 3 impurities characterized by different structures



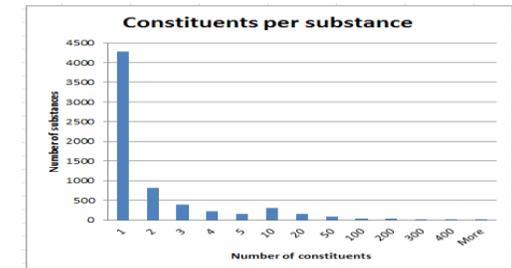
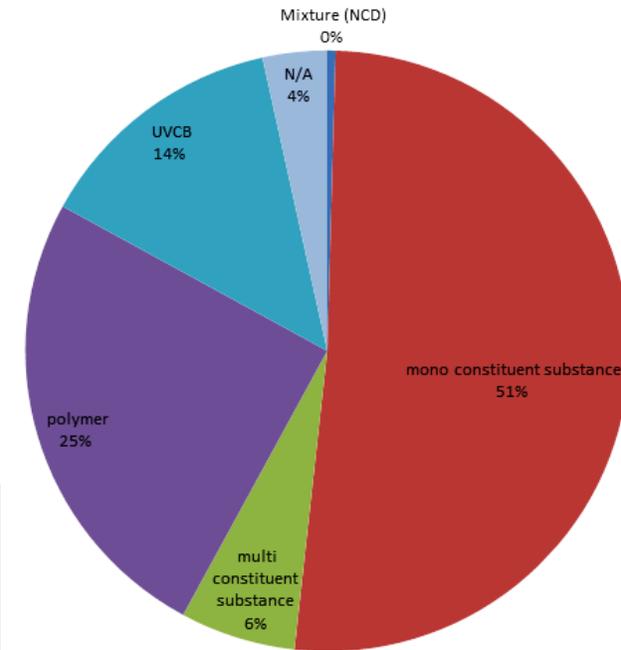
UIUCID interface showing substance definition parameters:

Constituents:  $\geq 80.0$ ,  $\leq 100.0$ , % (w/w)

Constituent Name	Chemical Name	EC Number	EC Number	Concentration Range	Unit
orthophosphoric Acid / phosphoric acid	7664-38-2 / 231-633-2	, $\geq 20.0$	$\leq 55.0$	% (w/w)	
methyl dihydrogen phosphate / methyl dihydrogen phosphate	812-00-0 / 212-379-1	, $\geq 25.0$	$\leq 75.0$	% (w/w)	
dimethyl hydrogen phosphate / dimethyl hydrogen phosphate	843-78-5 / 212-389-6	, $\geq 0.0$	$\leq 10.0$	% (w/w)	

Impurities:

Impurity Name	Chemical Name	EC Number	EC Number	Concentration Range	Unit
methanol / methanol	67-56-1 / 200-659-6	, $\geq 0.0$	$\leq 2.0$	% (w/w)	
pyro phosphoric acid, methyl esters / pyro phosphoric acid, methyl esters	, $\geq 0.0$	$\leq 3.0$	% (w/w)		
water / dihydrogen oxide	7732-18-5 / 231-791-2	, $\leq 15.0$	% (w/w)		



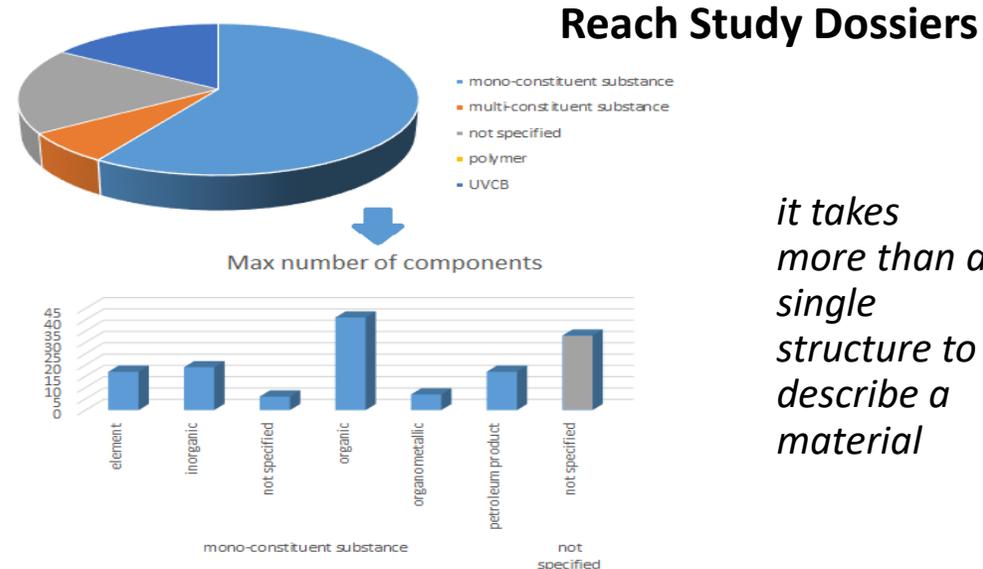
non-confidential REACH data  
 supplied by ECHA  
<http://ambitlri.ideaconsult.net>

# Chemical substance database (background)

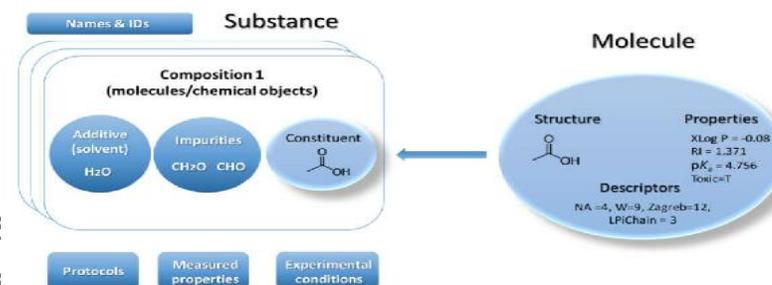
**ambit**

- The publicly accessible chemical databases generally associate chemical substances with a single structure and do not provide easily accessible data on substance identity and composition.
- Cefic's Long-Range Research Initiative (LRI) AMBIT software enable the representation of chemical substances in real industry conditions
- Open source database and web application <https://ambit.sourceforge.net>

*AMBIT Databases support substances (mono & multiconstituents, additives, impurities, UVCB), nanomaterials, advanced materials, microplastics*



*it takes more than a single structure to describe a material*



CHAPTER 3

**Cheminformatics Representation of Chemical Structures – A Milestone for Successful Big Data Modelling in Predictive Toxicology**

Nikolay Kochev, Nina Jeliaskova and Ivanka Tsakovska

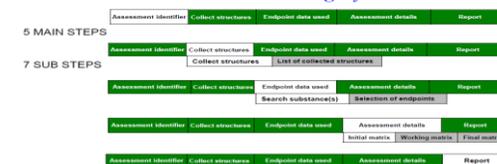




# LRI Read Across Tool

<https://ambitlri.ideaconsult.net>

Workflow for read across and category formation



ambit Read Across Assessment

Home New assessment New assessment from template Own assessments All assessments

Assessment identifier Collect structures (5) Select substances (0) Select endpoints (0) Assessment details (0) Report

Assessment title:  help Read Across workflow

Owner:  Workflow for read across and category formation. RACOS guidance

Purpose:  The assessment workflow is organized in five main tabs: Assessment identifier, Collect structures, Endpoint data used, Assessment details, Report

Version:  1

Version start date:  12.13.2022

Version last modified on:  12.13.2022

Status:  Draft Version

Assessment code:

Assessment (DocId)s:

Assessment ID:  a5225917-3438-4856-871f-5a5d464b02b3

Users with write access:

Users with read access:

Profile Assessment Create this content

ambit Read Across Assessment

Home New assessment New assessment from template Own assessments All assessments

Assessment identifier Collect structures (5) Select substances (0) Select endpoints (0) Assessment details (0) Report

Load structure:  Similarity:  Substructure:  Selected:

Showing from 1 to 5 in pages of 1 entries Previous list

Diagram	CASRN	EC number	Name	Risk status
	6890-37-4	79-89-4	isopropyl propanoate	target material
	67-63-0	200-961-7	isopropyl alcohol	Tier II read across analog
	16362-47-5	242-233-3	isopropyl decanoate	Tier II read across analog

ambit Read Across Assessment

Home New assessment New assessment from template Own assessments All assessments

Assessment identifier Collect structures (4) Select substances (3) Select endpoints (34) Assessment details (0) Report

Initial matrix Working matrix Final matrix

Structure:

Showing from 1 to 3 in pages of 1 entries Previous list

Diagram	CASRN	EC number	Name	Risk status
	67-63-0	200-961-7	isopropyl alcohol	target material
	16362-47-5	242-233-3	isopropyl decanoate	target material
	6890-37-4	79-89-4	isopropyl propanoate	target material

ambit Read Across Assessment

Home New assessment New assessment from template Own assessments All assessments

Assessment identifier Collect structures (5) Select substances (0) Select endpoints (0) Assessment details (0) Report

Initial matrix Working matrix Final matrix

Structure:

Showing from 3 to 4 in pages of 26 entries Previous list

Substance Name	Data source	Diagram	Chemical Name	4.7. Partition coefficient	2.2.1. Acute toxicity - oral
1-1 Dicyclopentadiene	Current Prokuba (Eubiochem) GmbH / Substanz am Team / Germany		5,5,8,11-Tetraazabicyclo[2.2.1]hept-2-ene	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)
1-2 Tetrahydrofuran	Current Prokuba (Eubiochem) GmbH / Substanz am Team / Germany		2,2,5,5-Tetrahydrofuran	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)
1-3 Nitrobenzene	Current Prokuba (Eubiochem) GmbH / Substanz am Team / Germany		1,1-Dinitroethane	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)
1-4 Nitrobenzene	Current Prokuba (Eubiochem) GmbH / Substanz am Team / Germany		1,1-Dinitroethane	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)

ambit Assessment Report

Create Word file

### Ambit Assessment Report

#### Glymes category Demo

Author: Unknown  
Date: 28.04.2019  
Assessment code: 276b33a-0017-4314-b682-3d25a5eade  
Purpose: Demonstration and testing

Assessment Identifiers

Assessment title: Glymes category Demo

Owner: Unknown

Date: 28.04.2019

Assessment code: 276b33a-0017-4314-b682-3d25a5eade

Purpose: Demonstration and testing

Version: 1

Status: published

Version start date: 28.04.2019

Version last modified on: 28.04.2019

Published: published

Assessment code: glydimo

Assessment DocId: local

Assessment ID: 276b33a-0017-4314-b682-3d25a5eade

The original assessment in Ambit can be found via Assessment ID

Showing from 1 to 4 in pages of 30 entries Previous list

Diagram	CASRN	EC number	Names	Rationale
	110-71-4	200-794-9	1,2-dimethoxyethane   Ethane, 1,2-dimethoxy-	same functional group, comparable metabolism
	111-96-6	200-824-4	1,2-dimethoxyethane   Ethane, 1,1'-oxybis(2-methoxy-   Ethane, 1,1'-oxybis(2-methoxy-	same functional group, comparable metabolism
	112-49-2	200-977-3	1,2-bis(2-methoxyethoxy)ethane   2,2,5,5,11-tetraoxadecane   2,5,8,11-tetraoxadecane	same functional group, comparable metabolism

Substance Name	CASRN	EC number	Assessment Endpoints	Assessment Results
1-1 Dicyclopentadiene	5,5,8,11-Tetraazabicyclo[2.2.1]hept-2-ene	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)
1-2 Tetrahydrofuran	2,2,5,5-Tetrahydrofuran	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)
1-3 Nitrobenzene	1,1-Dinitroethane	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)
1-4 Nitrobenzene	1,1-Dinitroethane	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)	0.333 (Temperature = 25.0 °C, pH = 7.0)

# Toxtree (toxtree.sf.net) Vega integration (collab with IRFMN)

Exact structure Similarity Substructure URL  Only hits with substance data 0.9 isopropyl propionate

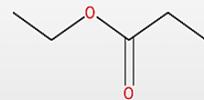
Refine the search within given dataset

Available structure attributes (1/6)	
CasRN	105-37-3
EC number	203-291-4
IUCLID 5 Reference substance UUID	IUC6-c1f0dc8a-6fbb-4f56-86e4-b52c03fd8c73
Names	Unnamed[ethyl propionate
SMILES	O=C(OCC)CC
Std. InChI key	FKRCODPIKNYEAC-UHFFFAOYSA-N
Std. InChI	InChI=1S/C5H10O2/c1-3-5(6)7-4-2/h3-4H2,1-2H3

VEGA models (0/1)

Model	Prediction
Hepatotoxicity model (IRFMN)	Unknown
Skin Sensitization model (CAESAR)	Sensitizer
Developmental Toxicity model (CAESAR)	Toxicant
Thyroid Receptor Beta effect (NRMEA)	Inactive
Ready Biodegradability model (IRFMN)	Possible Readily Biodegradable
Androgen Receptor-mediated effect (IRFMN-COMPARA)	NON-active
Estrogen Receptor Relative Binding Affinity model (IRFMN)	Inactive
Estrogen Receptor-mediated effect (IRFMN-CERAPP)	Possible NON-active
In vitro Micronucleus activity (IRFMN-VERMEER)	Not predicted
Mutagenicity (Ames test) model (CAESAR)	NON-Mutagenic
Carcinogenicity model (CAESAR)	Carcinogen
Mutagenicity (Ames test) model (SarPy-IRFMN)	Possible NON-Mutagenic
Thyroid Receptor Alpha effect (NRMEA)	Inactive

Structure diagram



Adipose tissue - blood model (INERIS) - [assessment] : 0.389 (MODERATE reliability) null  
Adipose tissue - blood model (INERIS) - [prediction] : 0.389 Log units  
Adipose tissue - blood model (INERIS) - [reliability] : MODERATE reliability null

Androgen Receptor-mediated effect (IRFMN-COMPARA) - [assessment] : NON-active (EXPERIMENTAL value) null  
Androgen Receptor-mediated effect (IRFMN-COMPARA) - [experimental value] : NON-active null  
Androgen Receptor-mediated effect (IRFMN-COMPARA) - [reliability] : GOOD reliability null

BCF model (CAESAR) - [assessment] : 0.46 log(L/kg) (MODERATE reliability) null  
BCF model (CAESAR) - [prediction] : 0.456 log(L/kg)  
BCF model (CAESAR) - [reliability] : MODERATE reliability null

Carcinogenicity model (CAESAR) - [assessment] : Carcinogen (MODERATE reliability) null  
Carcinogenicity model (CAESAR) - [reliability] : MODERATE reliability null

Daphnia Magna LC50 48h (EPA) - [assessment] : 147.65 mg/L (EXPERIMENTAL value) null

## toxtree models

all modules (3.1.0)

## vega models

Mutagenicity (Ames test) model (CAESAR) (2.1.14)

Estrogen Receptor-mediated effect (IRFMN-CERAPP) (1.0.1)

Estrogen Receptor Relative Binding Affinity model (IRFMN) (1.0.2)

Thyroid Receptor Beta effect (NRMEA) (1.0.1)

Ready Biodegradability model (IRFMN) (1.0.10)

Total body elimination half-life (QSARINS) (1.0.1)

Fathead Minnow LC50 96h (EPA) (1.0.10)

Adipose tissue - blood model (INERIS) (1.0.1)

Developmental Toxicity model (CAESAR) (2.1.8)

Skin Permeation (LogKp) model (Ten Berge) (1.0.1)

Androgen Receptor-mediated effect (IRFMN-COMPARA) (1.0.1)

Skin Permeation (LogKp) model (Potts and Guy) (1.0.1)

BCF model (CAESAR) (2.1.15)

Mutagenicity (Ames test) model (SarPy-IRFMN) (1.0.8)

Thyroid Receptor Alpha effect (NRMEA) (1.0.1)

Carcinogenicity model (CAESAR) (2.1.10)

Hepatotoxicity model (IRFMN) (1.0.1)

Daphnia Magna LC50 48h (EPA) (1.0.9)

In vitro Micronucleus activity (IRFMN-VERMEER) (1.0.1)

Skin Sensitization model (CAESAR) (2.1.7)

# LRI Read Across Tool Assessment report with final matrix (section 5 & 6)

## 5. Assessment data matrix

In the following, for each substance, the associated endpoint data are given, either experimental data, waiving or read-across.

For detailed data or rationale for waiving and read-across, click hyperlinks in the table. These data or rationales can also be found in the annex of the report.

Tag	<a href="#">CM</a> <a href="#">CM</a>	<a href="#">CM</a> <a href="#">CM</a> <a href="#">CM</a>	
Substance name	Diglyme	Tetraglyme	
CAS No.			
4.7. Partition coefficient	<a href="#">-0.36</a> (Temperature = 25.0 °C, pH = 7.0) <a href="#">0</a> <a href="#">-0.4</a> (Temperature = , pH = ) <a href="#">0</a>	<a href="#">-0.84</a> (Temperature = 23.0 °C, pH = ca.7.0 ca. ) <a href="#">0</a>	
7.2.1. Acute toxicity - oral	<a href="#">LD50 = 4760 mg/kg bw</a> (Species = rat) <a href="#">0</a>	<a href="#">LD50 = 3850 mg/kg bw</a> (Species = rat) <a href="#">0</a>	<a href="#">LD50</a> <a href="#">0</a>
7.2.2. Acute toxicity - inhalation	<a href="#">LC0 = 11 mg/L air</a> (Species = rat) <a href="#">0</a>		
7.2.3. Acute toxicity - dermal			
7.3.1. Skin irritation / Corrosion	<a href="#">erythema score = 0.89</a> <a href="#">0</a> <a href="#">edema score = 0.5</a> <a href="#">0</a>	<a href="#">erythema score = 0.8</a> <a href="#">0</a> <a href="#">edema score = 0</a> <a href="#">0</a> <a href="#">edema score = 0</a> <a href="#">0</a> <a href="#">erythema score = 1.3</a> <a href="#">0</a> <a href="#">erythema score = 1</a> <a href="#">0</a> <a href="#">edema score = 0</a> <a href="#">0</a> <a href="#">edema score = 0</a> <a href="#">0</a> <a href="#">erythema score = 0</a> <a href="#">0</a>	
7.3.2. Eye irritation	<a href="#">Maximum mean total score (MMTS) = 13</a> <a href="#">0</a> <a href="#">cornea score = 0</a> <a href="#">0</a> <a href="#">iris score = 0</a> <a href="#">0</a> <a href="#">conjunctivae score = 0.78</a> <a href="#">0</a> <a href="#">chemosis score = 0.28</a> <a href="#">0</a>	<a href="#">chemosis score = 0</a> <a href="#">0</a> <a href="#">chemosis score = 0</a> <a href="#">0</a> <a href="#">iris score = 0</a> <a href="#">0</a> <a href="#">cornea score = 0</a> <a href="#">0</a> <a href="#">cornea score = 0</a> <a href="#">0</a> <a href="#">iris score = 0</a> <a href="#">0</a> <a href="#">chemosis score = 0</a> <a href="#">0</a> <a href="#">conjunctivae score = 0.3</a> <a href="#">0</a> <a href="#">cornea score = 0</a> <a href="#">0</a> <a href="#">iris score = 0</a> <a href="#">0</a> <a href="#">conjunctivae score = 0</a> <a href="#">0</a> <a href="#">chemosis score = 0</a> <a href="#">0</a> <a href="#">cornea score = 0</a> <a href="#">0</a> <a href="#">conjunctivae score = 0.1</a> <a href="#">0</a> <a href="#">iris score = 0</a> <a href="#">0</a> <a href="#">conjunctivae score = 0</a> <a href="#">0</a>	
7.5.1. Repeated dose toxicity - oral		<a href="#">NOEL = 250 mg/kg bw/day (actual dose received)</a> (Species = rat, Test type = subacute) <a href="#">0</a>	
7.5.2. Repeated dose toxicity - inhalation	<a href="#">NOEC = 370 ppm</a> (Species = rat, Test type = subacute) <a href="#">0</a> <a href="#">NOAEC = 110 ppm</a> (Species = rat, Test type = subacute) <a href="#">0</a>		
7.6.1. Genetic toxicity in vitro	<a href="#">negative</a> (Study type = bacterial reverse mutation assay (e.g. Ames test), OECD)		

➤ The report describes the read across workflow in 6 sections and 5 annexes

## 6. Justification for read-across / category

### Annex 1 Experimental data

[Create Excel file with all used experimental data](#)

### Annex 2 Rationale for gap filling

#### Substance 3: Monoglyme

7.2.1. Acute toxicity - oral

Endpoint	Value	Conditions	Guideline or Justification
LD50	3850 mg/kg bw/day	-	

Rationale for read-across based on grouping of substances (category approach)  
Justified

### Annex 3 Rationale for deleting experimental data

#### Substance 4: Triglyme

7.2.1. Acute toxicity - oral

Endpoint	Value	Conditions		Guideline or Justification
		Sex	Species	
LD50	5877 mg/kg bw	female	rat	OECD Guideline 401 (Acute Oral Toxicity)

lower value exists

### Annex 4 Initial matrix

[Create Excel file with the initial matrix](#)

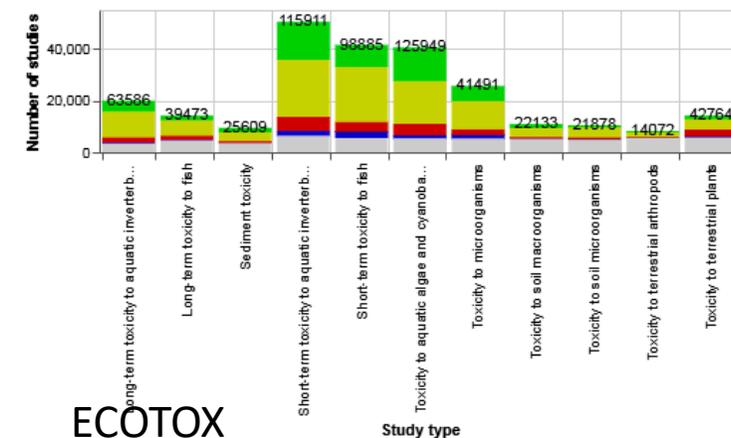
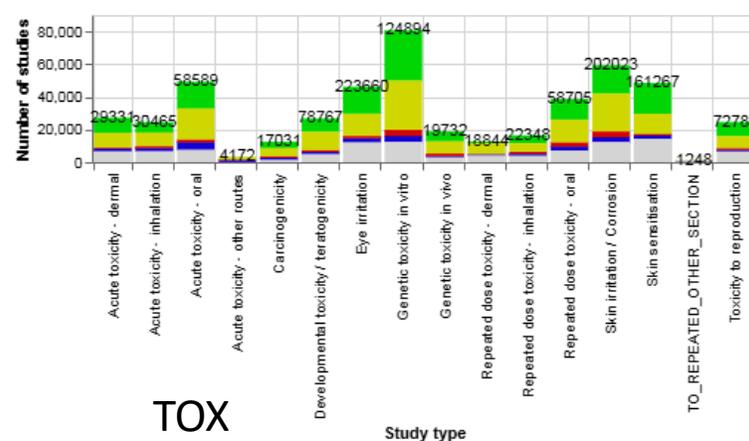
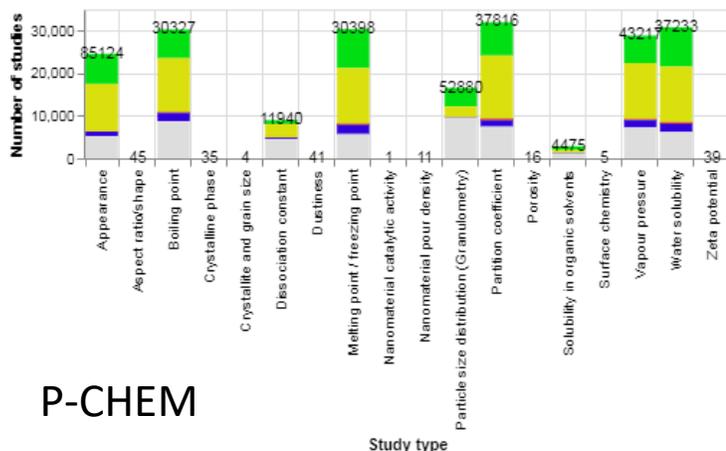
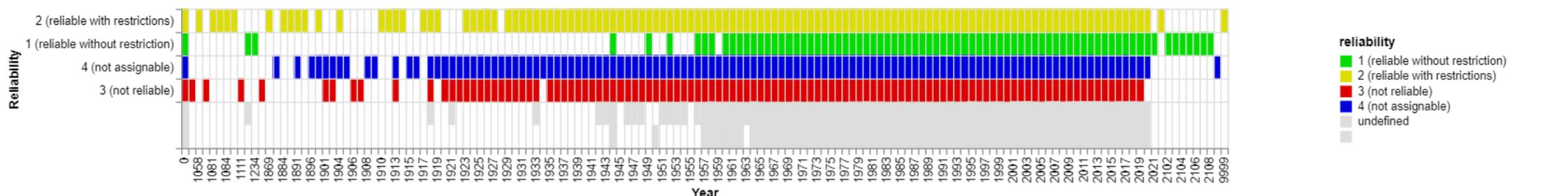
### Annex 5 Working matrix

[Create Excel file with the working matrix](#)

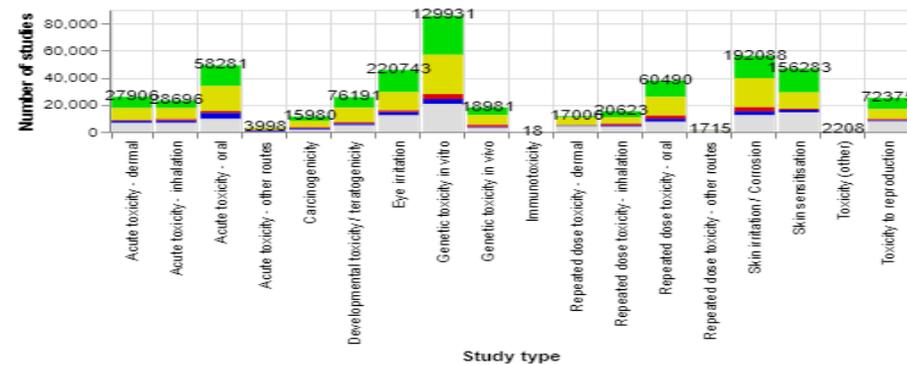
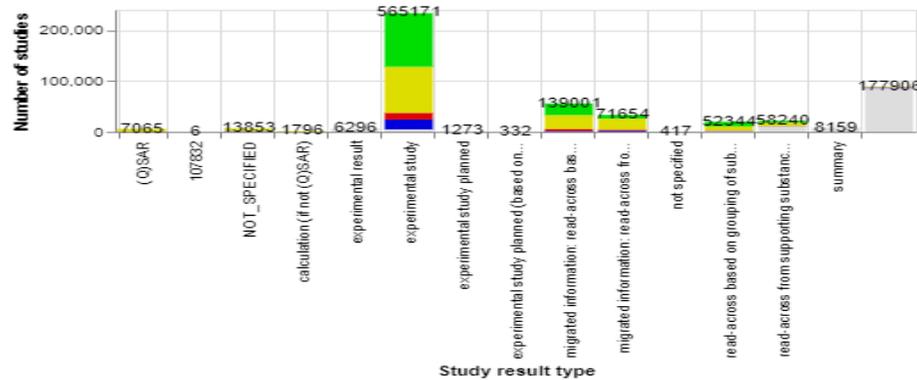
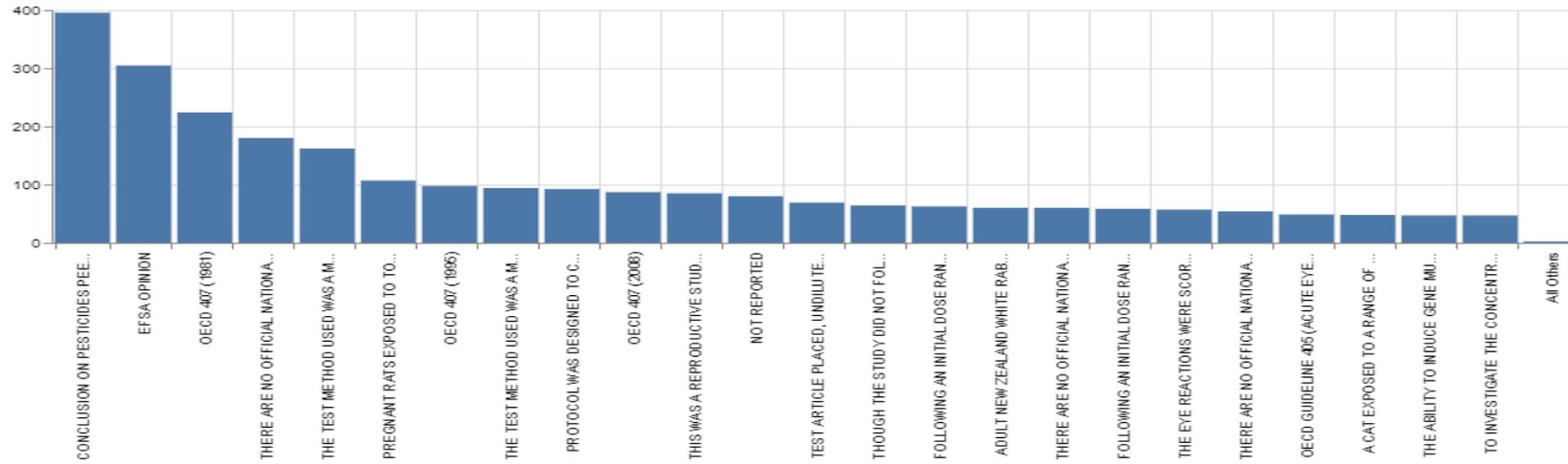
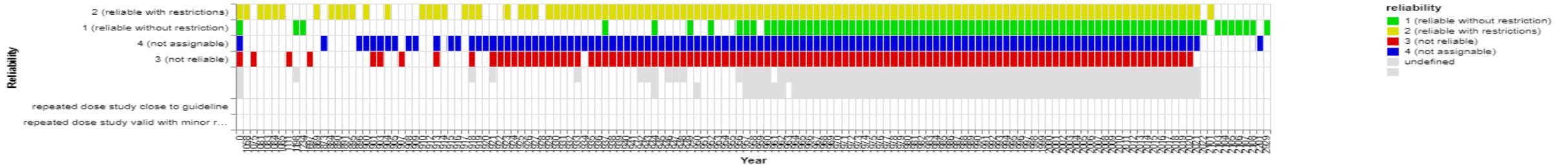
# Dashboards : Data quality (REACH dossiers)

## Summary

### P-CHEM

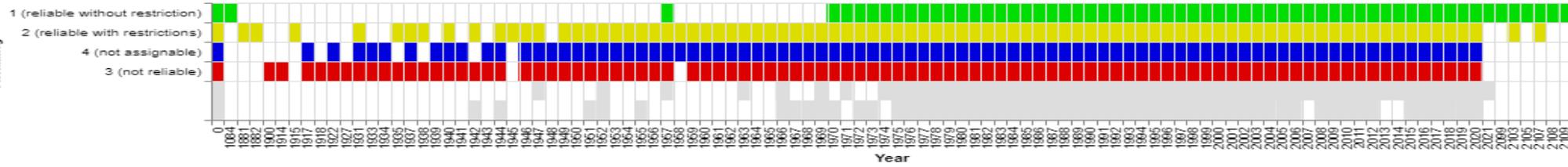


# Dashboards : TOX Data quality (REACH dossiers)



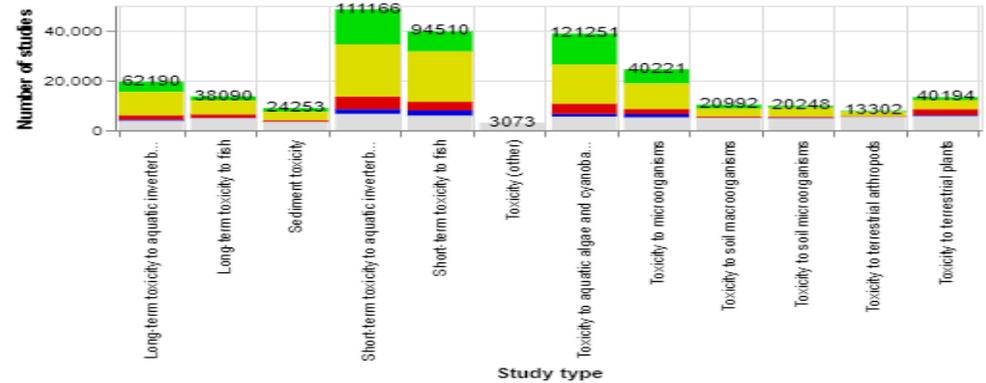
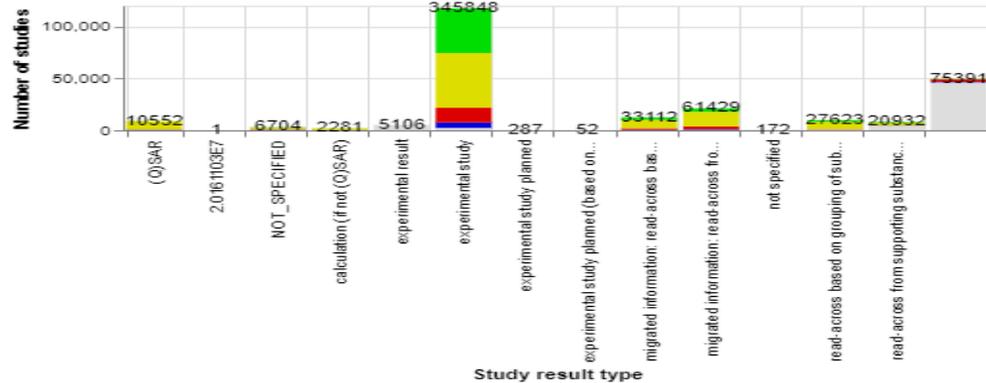
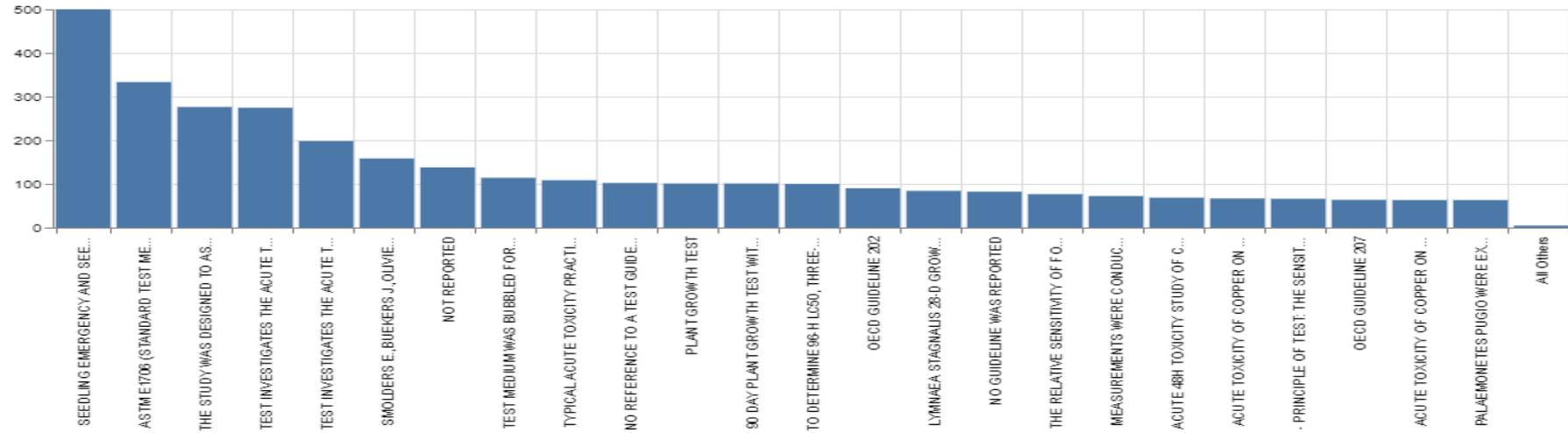
# Dashboards : ECOTOX Data quality (REACH dossiers)

Reliability



reliability

- 1 (reliable without restriction)
- 2 (reliable with restrictions)
- 3 (not reliable)
- 4 (not assignable)
- undefined



# AMBIT Search for Structures & Endpoint data

ambit cefic LRD The Long-term Research Initiative

Search Assessments Import Enhanced functions Admin Help [ambit3] Log out

Search structures and associated data

Exact structure Similarity Substructure URL  Enable fragment search

Identifiers Datasets **1) Find Structure(s)** Export

Showing from 1 to 1 in pages of 20 entries Previous Next

Diagram	CasRN	EC number	IUCLID 5 R	Names	Trade Name	IUPAC name	SMILES	Std. InChI key	Std. InChI	REACH registration date	Similarity
	111-96-6	203-924-4	ECHA-d1...	1-methoxy-2-(2-methoxyethoxy)ethane, bis(2-methoxyethyl) ether Ethane, 1,1'-oxybis 2-methoxy- Ethane, 1,1'-oxybis[2-methoxy-	Diglyme	-	O(C)CCO...	SBZXBUIDTXKZTM-UHFFFAOYSA-N	InChI=1S...	30.11.2010	-

Select data sources/models

2) Find Substance(s)

Showing from 1 to 8 in pages of 20 substances Previous Next

Substance Name	Substance UUID	Substance Type	Public name	Reference substance UUID	Owner	Info	Contained in as
bis(2-methoxyethyl) ether	ECHA-719..	mono constituent substance	Diglyme	ECHA-eab..			constituent
Glymes Diethylene glycol dimethyl ether (DEGDME, Diglyme) ICS MSe DE71	IUC5-a43...	mono constituent substance	Diglyme	ECB5-196...	Clariant Produkte (Deutschland) GmbH / Sulzbach am Taunus / Germany		constituent

IUC Substance Composition Tox (12) P-Chem (4) Eco Tox (8) Env Fate (4)

Composition name: Crude Diglyme  
Composition UUID: L-327712f8-626c-4bd0-b915-621c4b5f87c0  
Purity of IUC Substance: 85-99 % (w/w)

Type	Name	EC No.	CAS No.	Typical concentration	Concentration ranges	
Constituent	1-Methoxy-2-(2-Methoxyethoxy)Ethane,Bis(2-Methoxyethyl) Ether,Ethane, 1,1'-Oxybis 2-Methoxy-,Ethane, 1,1'-Oxybis[2-Methoxy-	203-924-4	111-96-6	ca.92 % (w/w)	>=85 % (w/w) <=99 % (w/w)	Also contained in...
Impurity	2-(2-Methoxyethoxy)Ethanol,Diethyleneglycolmonomethyl_ether,Ethanol, 2-(2-Methoxyethoxy)-,2-(2-Methoxyethoxy)Ethanol,2-(2-Methoxyethoxy)Ethan-1-ol,Mdg,Diethylene Glycol Methyl Ether,3,6-Dioxa-1-Heptanol,2-Hydroxyethyl 2-Methoxyethyl Ether	203-906-6	111-77-3	ca.8 % (w/w)	>=1 % (w/w) <=15 % (w/w)	Also contained in...
Impurity	Unknown,Unknown Organic Impurities			<1 % (w/w)	0 % (w/w) <1.5 % (w/w)	Also contained in...

### 3) Display data

IUC Substance Composition Tox (12) P-Chem (4) Eco Tox (8) Env Fate (4)

Filter...

Diglyme

6.1.1 Short-term toxicity to fish (1)

Test Medium	Organism	Meas. Conc.	Expos	Endpoint	Effect	Based on	Guideline	Owner	Reliability	UUID
freshwater	Leuciscus idus	meas. (not spc)	96 h	LC0	> 2000 mg/L	mortality	test mat: DIN 38412, L15	Clariant	2 (reliable with restrictions)	IUC5-20...
		meas. (not spc)	96 h	NOEC	2000 mg/L	mortality	test mat: L15			

Showing 1 study(s) (1 to 1)

6.1.3 Short-term toxicity to aquatic invertebrates (2)

6.1.4 Long-term toxicity to aquatic invertebrates (1)

6.1.5 Toxicity to aquatic algae and cyanobacteria (2)

6.1.7 Toxicity to microorganisms (2)

# AMBIT Search for Structures

Exact, Similarity & Substructure

ambit cefic LRD

Search ▾ Assessments ▾ Import ▾ Enhanced functions ▾ Admin ▾ Help ▾ [Log in](#)

Search structures and associated data

Exact structure Similarity Substructure URL  Enable fragment search

Identifiers Datasets Export

Showing from 1 to 1 in pages of 20 entries [Previous](#) [Next](#) Filter...

	Diagram	CasRN	EC number	IUCLID 5 R <sub>i</sub>	Names	Trade Name	IUPAC name	SMILES	Std. InChI key	Std. InChI	REACH registration date
- 1 -		111-96-6	203-924-4	ECHA-d1...	1-methoxy-2-(2-methoxyethoxy)ethane, bis(2-methoxyethyl) ether[Ethane, 1,1'-oxybis 2-methoxy- Ethane, 1,1'-oxybis[2-methoxy-	Diglyme	-	O(C)CCO...	SBZXBUIDTXKZTM-UHFFFAOYSA-N	InChI=1S...	30.11.2010

Exact structure Similarity Substructure URL  Only hits with substance data 0.9

Identifiers Datasets Export

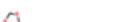
Showing from 1 to 3 in pages of 20 entries [Previous](#) [Next](#) Filter...

	Diagram	CasRN	IUCLID 5 Reference subst	Names	Similarity
- 2 -		111-46-6	ECHA-d6...	2,2'-oxydiethanol[2,2'-oxybisethanol Ethanol, 2,2'-oxybis-[dihydroxy diethyl ether[2,2'-oxybisethanol Diethylene glycol, 2,2'-oxydiethanol[(2-hydroxyethoxy)ethan-2-ol[2-(2-hydroxyethoxy)ethanol	1
- 3 -		68909-76-2	-	Ethanol, 2,2'-oxybis-, reaction products with ammonia, fractionation forecuts	1
- 1 -		110-80-5	ECHA-50...	2-ethoxyethanol[Ethanol, 2-ethoxy- Monoethyleneglycol ethyl ether[Ethylglycol EEMEG, Ethylene glycol monoethyl ether	0.92

Exact structure Similarity Substructure URL Custom SMARTS

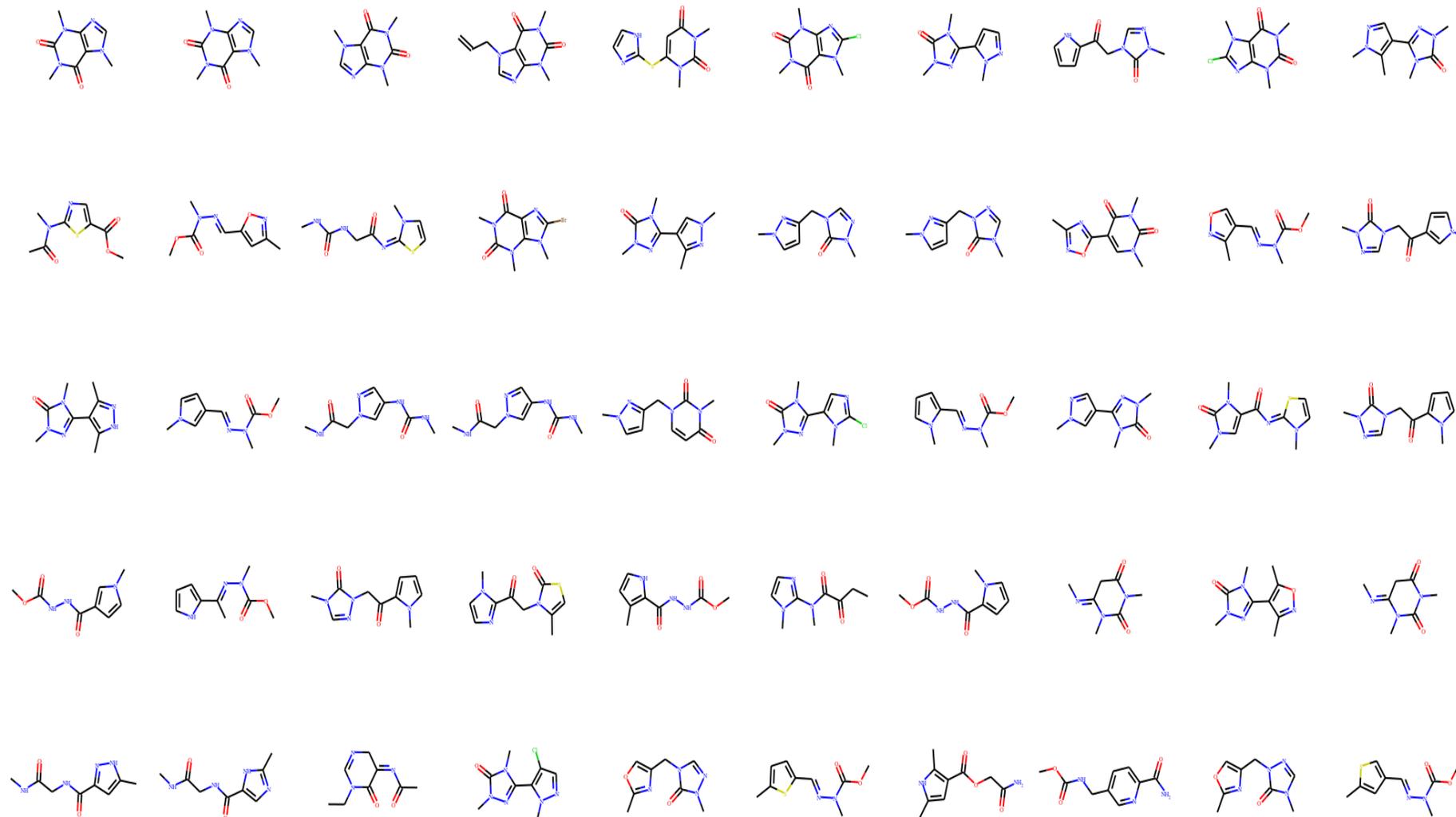
Identifiers Datasets

Showing from 1 to 20 in pages of 20 entries [Previous](#) [Next](#)

	Diagram	IUCLID 5 Reference subst	Names	Std. InChI key
- 1 -		-		HMFHBZSHGGEWLO-SOOFDHNKSA-N
- 2 -		-		CIWBSHSHKDKBQ-JLAZNSOCSA-N
- 3 -		-		FIPWRUSWJWJAL-UHFFFAOYSA-N

# Structure similarity search (>300 mln structures)

2018-12-19 12:15:15,900 INFO Query completed in 1.247268e-01 s



The screenshot shows the ExCAPE-DB web interface. The search results are displayed in a table with columns for 'Hits list' and 'Selection'. The results include the following entries:

Gene symbol (998131)	Hit	Selection
ADAT	ADAT	ADAT
ADAT1	ADAT1	ADAT1
ADAT2	ADAT2	ADAT2
ADAT3	ADAT3	ADAT3
ADAT4	ADAT4	ADAT4
ADAT5	ADAT5	ADAT5
ADAT6	ADAT6	ADAT6
ADAT7	ADAT7	ADAT7
ADAT8	ADAT8	ADAT8
ADAT9	ADAT9	ADAT9
ADAT10	ADAT10	ADAT10
ADAT11	ADAT11	ADAT11
ADAT12	ADAT12	ADAT12
ADAT13	ADAT13	ADAT13
ADAT14	ADAT14	ADAT14
ADAT15	ADAT15	ADAT15
ADAT16	ADAT16	ADAT16
ADAT17	ADAT17	ADAT17
ADAT18	ADAT18	ADAT18
ADAT19	ADAT19	ADAT19
ADAT20	ADAT20	ADAT20
ADAT21	ADAT21	ADAT21
ADAT22	ADAT22	ADAT22
ADAT23	ADAT23	ADAT23
ADAT24	ADAT24	ADAT24
ADAT25	ADAT25	ADAT25
ADAT26	ADAT26	ADAT26
ADAT27	ADAT27	ADAT27
ADAT28	ADAT28	ADAT28
ADAT29	ADAT29	ADAT29
ADAT30	ADAT30	ADAT30
ADAT31	ADAT31	ADAT31
ADAT32	ADAT32	ADAT32
ADAT33	ADAT33	ADAT33
ADAT34	ADAT34	ADAT34
ADAT35	ADAT35	ADAT35
ADAT36	ADAT36	ADAT36
ADAT37	ADAT37	ADAT37
ADAT38	ADAT38	ADAT38
ADAT39	ADAT39	ADAT39
ADAT40	ADAT40	ADAT40

## Industry-scale application and evaluation of deep learning for drug target prediction

Noé Sturm, Andreas Mayr, Thanh Le Van, Vladimir Chupakhin, Hugo Ceulemans, Joerg Wegner, Jose-Felipe Golib-Dzib, Nina Jeliakova, Yves Vandriessche, Stanislav Böhm, Vojtech Cima, Jan Martinovic, Nigel Greene, Tom Vander Aa, Thomas J. Ashby, Sepp Hochreiter, Ola Engkvist, Günter Klambauer & Hongming Chen

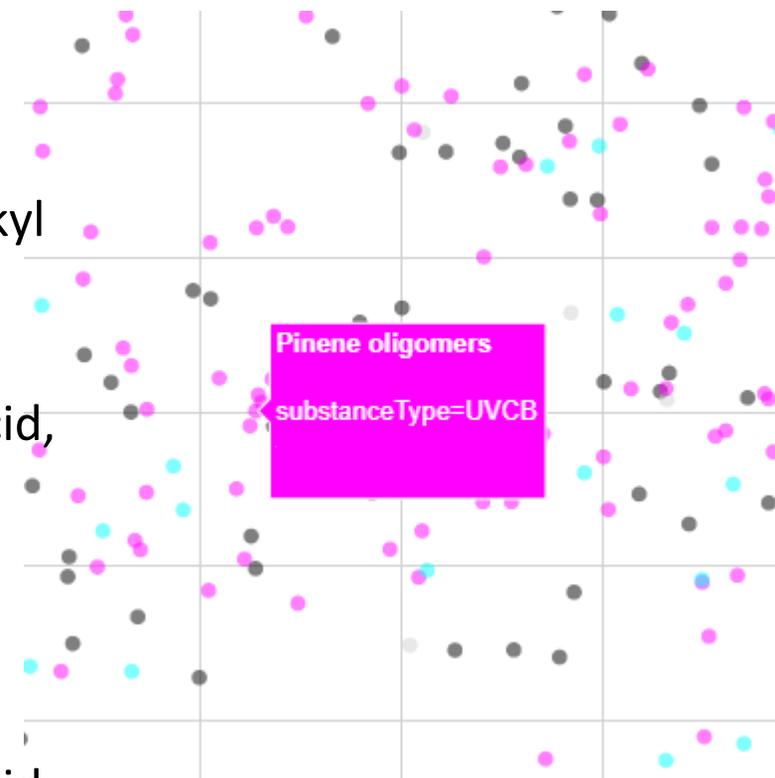
Journal of Cheminformatics 12, Article number: 26 (2020) | Cite this article

# AMBIT3 : Substance similarity query (example)

- Query: Pinene oligomers (UVCB)

- Hits:

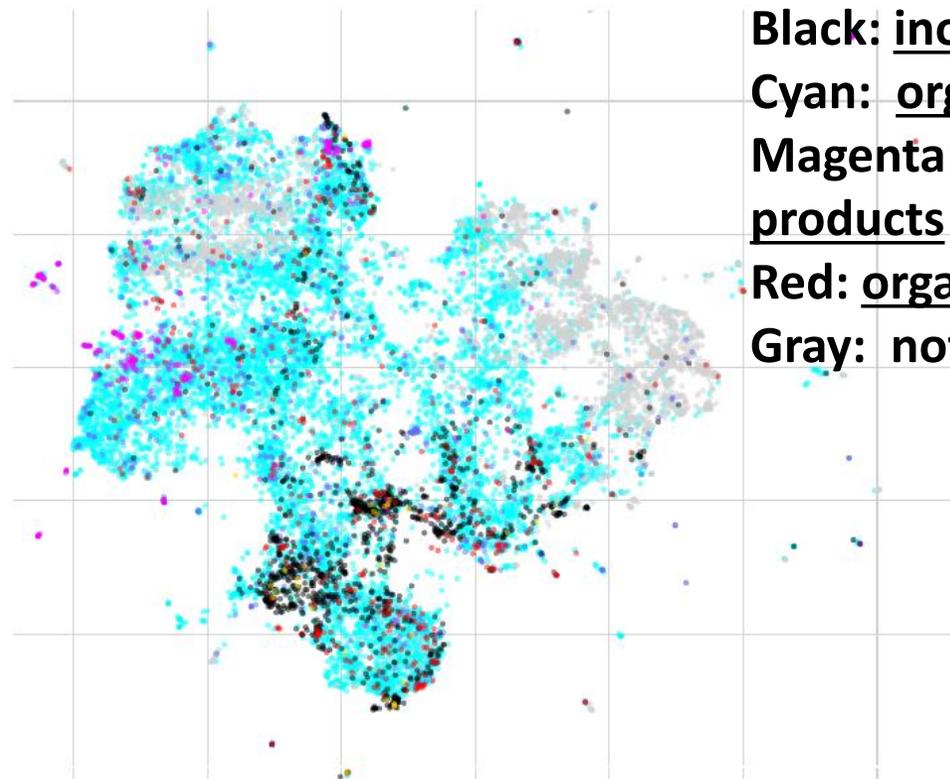
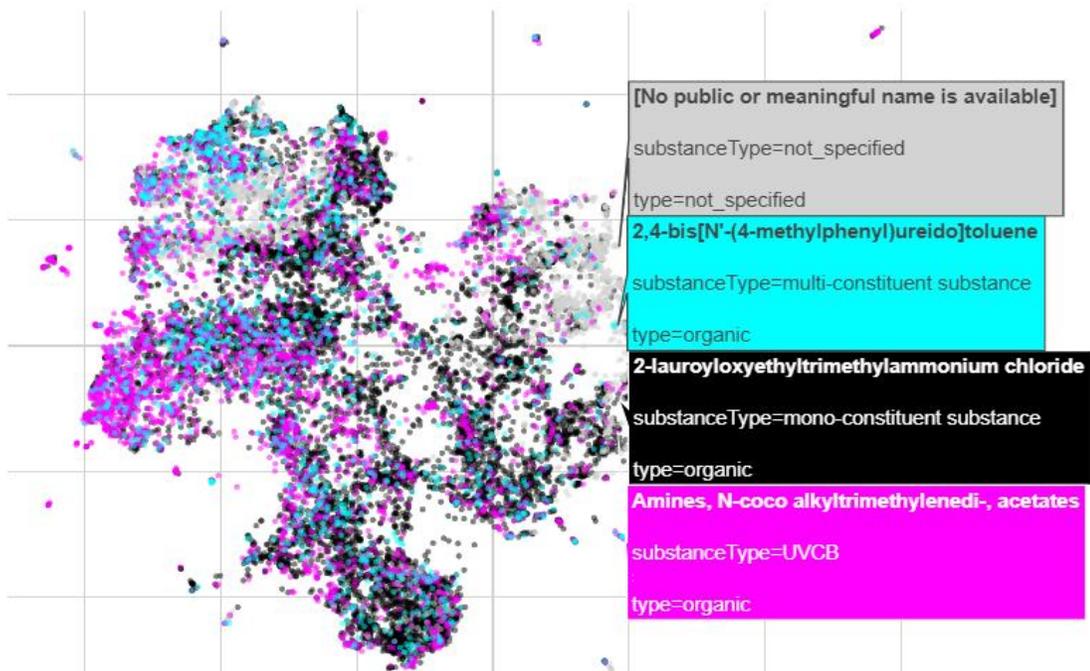
- Benzenesulfonic acid, 4-mono-C20-24 (even numbered)-alkyl derivs., magnesium salts Benzenesulfonic acid, mono-C20-24-alkyl derivs., magnesium salts
- Tetradecyl methacrylate 2-Propenoic acid, 2-methyl-, tetradecyl ester
- Decanoic acid, mixed esters with dipentaerythritol, heptanoic acid, isononanoic acid, and octanoic acid
- Benzenesulfonic acid, 4-(branched alkyl derivs.) and benzenesulfonic acid, 4-(linear alkyl derivs.), calcium salts
- Benzenesulfonic acid, C14-44-branched and linear alkyl derivs., calcium salts, overbased
- (Z)-9-Octadecen-1-ol ethoxylated
- Decanoic acid, mixed esters with dipentaerythritol, heptanoic acid and octanoic acid. Hatcol 5127



# AMBIT 3 Data maps (REACH dossiers)

• mono-constituent substance • not\_specified • UVCB • multi-constituent substance  
• polymer

• organic • not\_specified • inorganic • organometallic • petroleum product  
• element

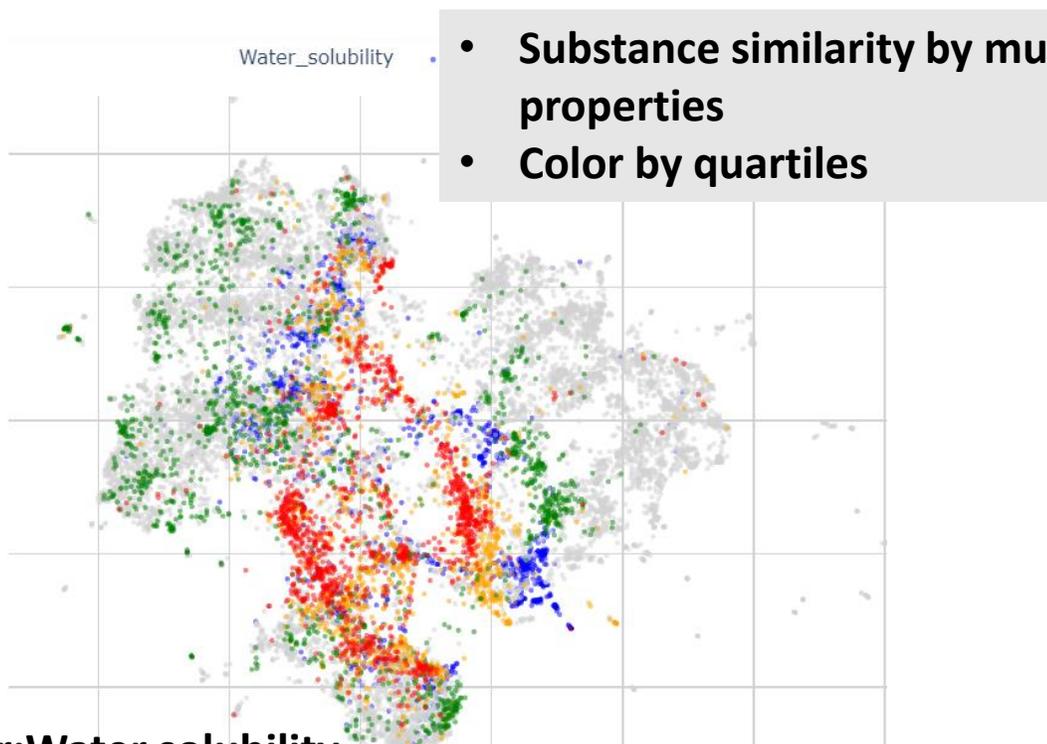


**Black: inorganic**  
**Cyan: organic**  
**Magenta: petroleum products**  
**Red: organometallic**  
**Gray: not specified**

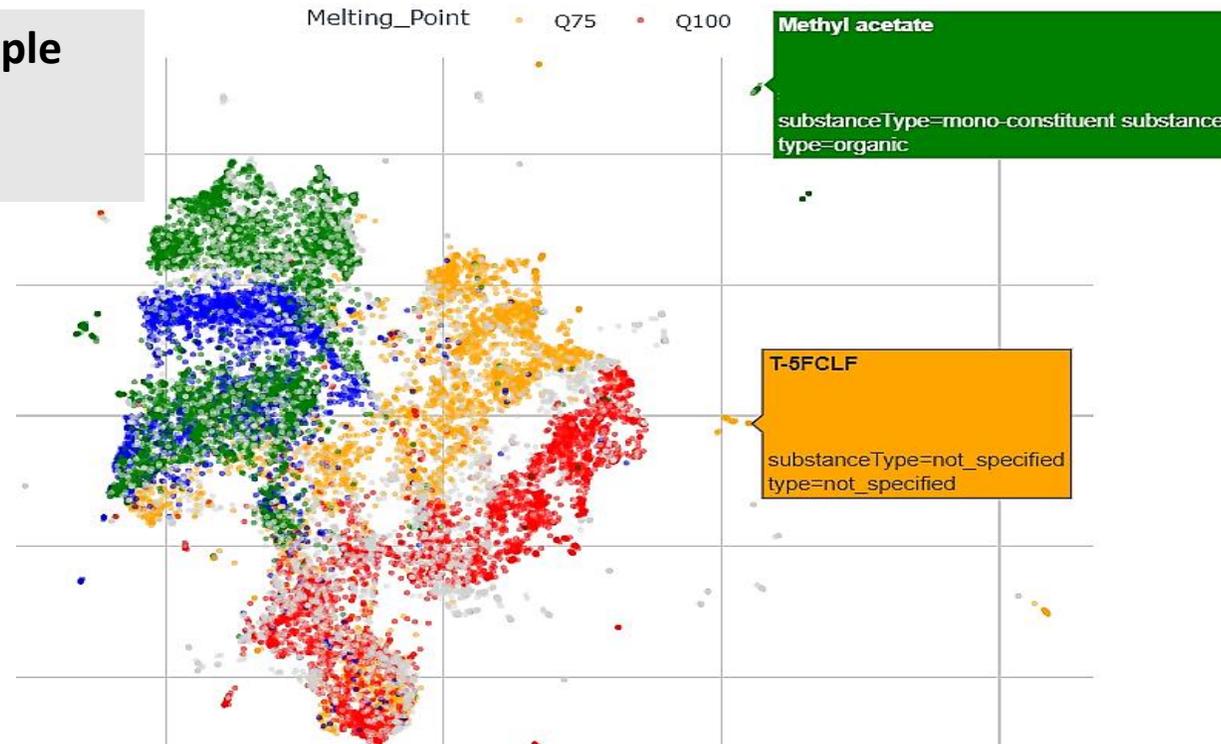
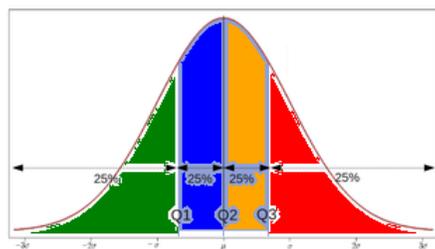
**Black: mono-constituent**  
**Cyan: multi-constituent**  
**Magenta: UVCB**  
**Gray: not specified**

- Substance similarity by multiple properties
- Similar substances = close points

# AMBIT 3 Data maps (REACH dossiers)



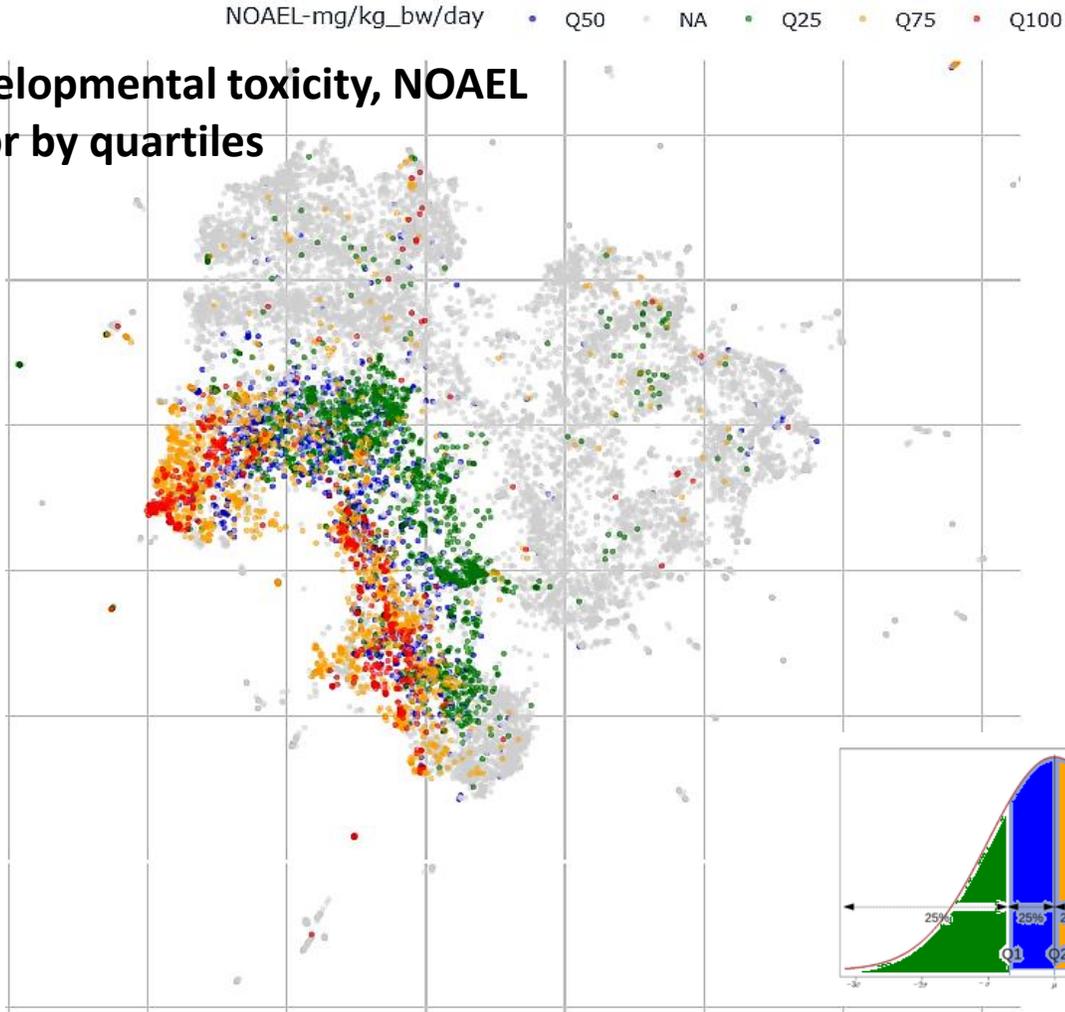
**Color: Water solubility**  
Green : low  
Blue : below median  
Orange : above median  
Red : highest water solubility



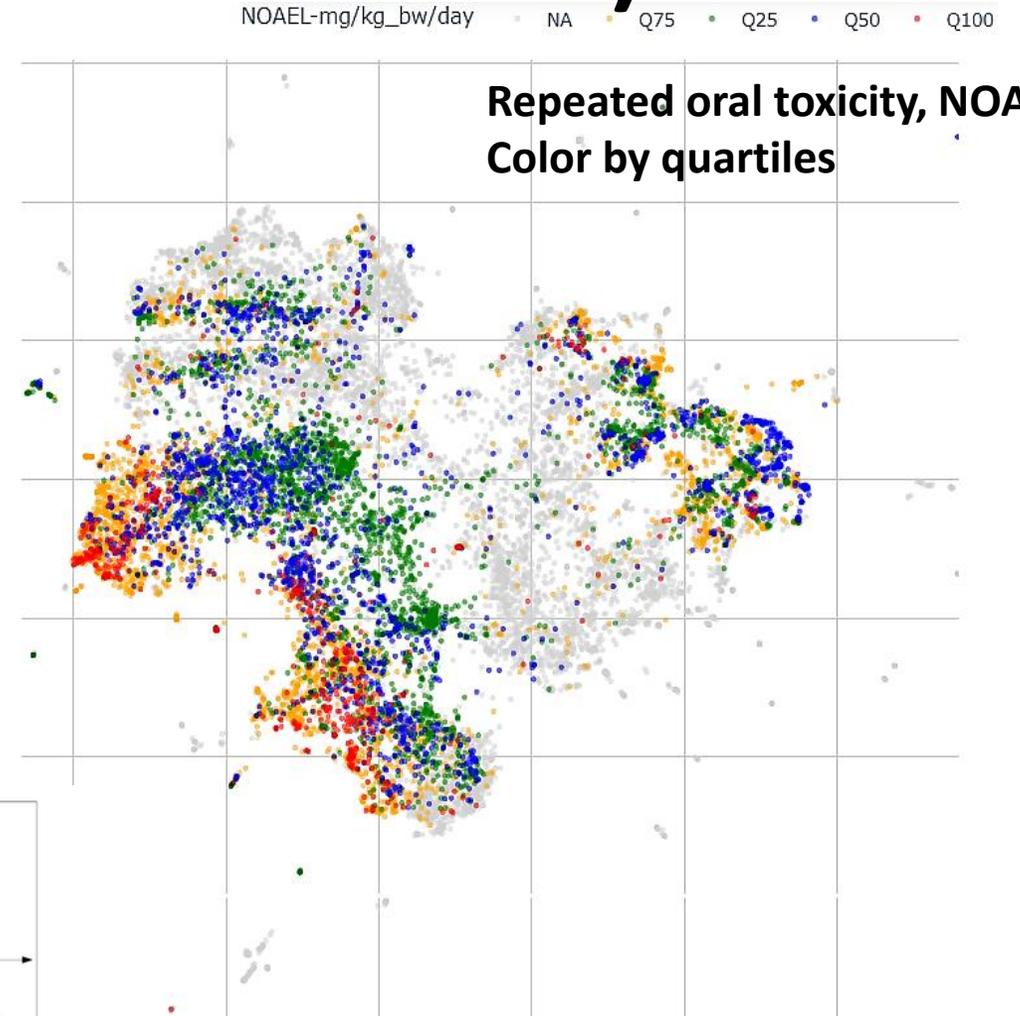
**Color: Melting point**  
Green : low  
Blue : below median  
Orange : above median  
Red : highest melting point

# AMBIT 3 Data maps (REACH dossiers)

Developmental toxicity, NOAEL  
Color by quartiles



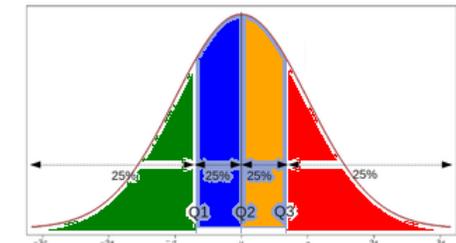
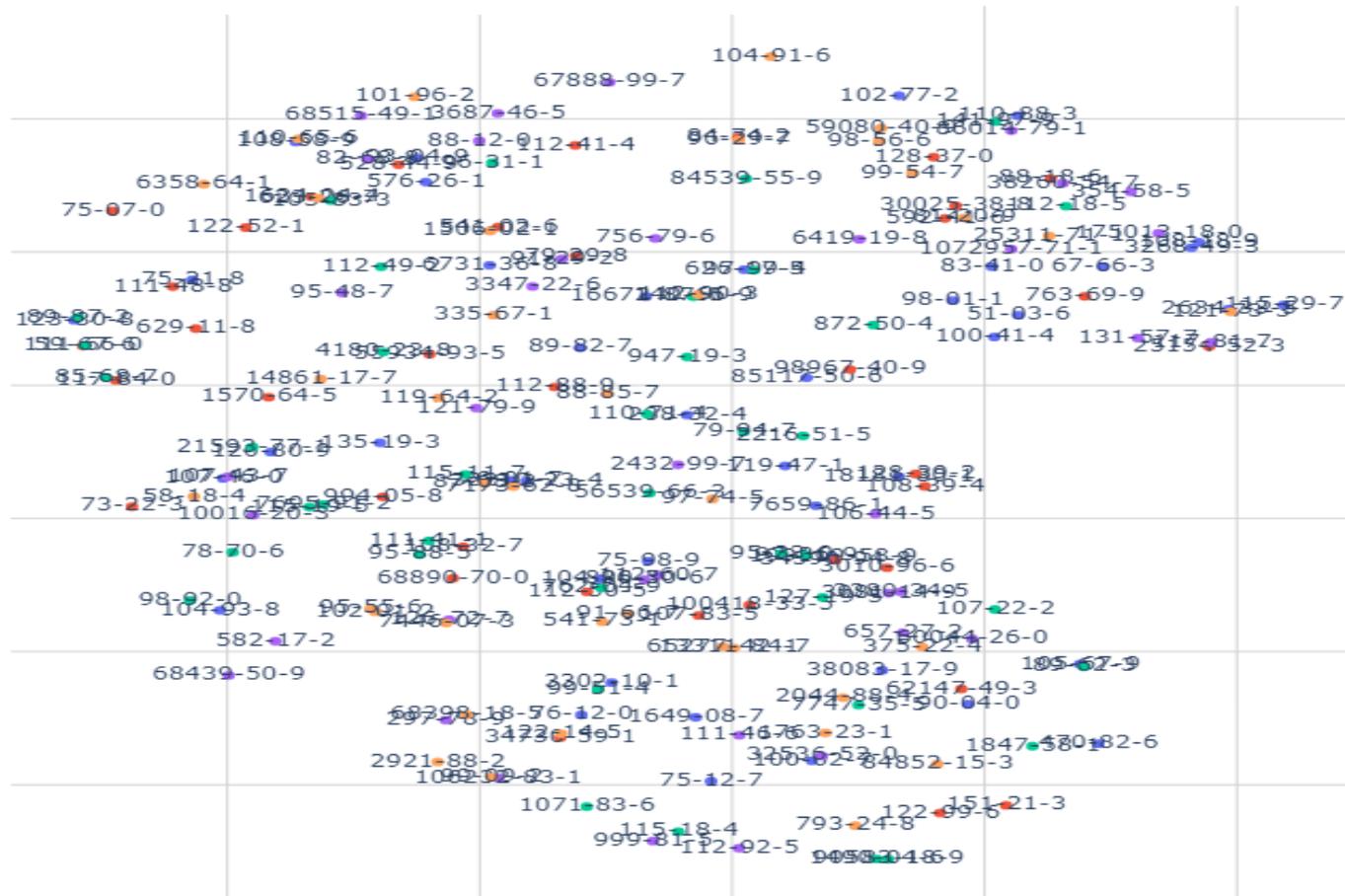
Repeated oral toxicity, NOAEL  
Color by quartiles



# AMBIT 3 Data maps (REPDOSE , Fraunhofer ITEM)

TO\_REPEATED\_ORAL\_SECTION LOEL (mg/kg bw/d) • Q50 • Q100 • Q75 • NA • Q25

Repeated oral toxicity, LOEL  
Color by quartiles



# Substance similarity search, based on the maps (User Interface)

The screenshot displays the 'Substance search' web application interface. The browser address bar shows the URL `ambitri-dev.ideaconsult.net/app/tool3/eda/index.html`. The application header includes the 'ambit3 Substance search' logo and navigation links for Home, Search, Read across, and Predict. A search bar is present in the header.

The main search area contains a search input field with the text 'PFOA', a page number input field set to '0', and a dropdown menu for search type set to 'simple'. Below the search input, the selected substance is identified as 'PFOA 335-67-1'. A table lists search results, with the first row highlighted in green and a red box around the 'Similarity type' column.

Similarity type	Page	Number of hits
substance PFOA 335-67-1	0	20
mono-constituent substance/organic (0.46) 2,2'-methylenediphenyl diisocya...		
mono-constituent substance/organic (0.44) 5-acetylamino-N,N'-bis(2,3-dihy...		
mono-constituent substance/organic (0.44) 4-methyl-4'-n-propyl-[1,1-biphe...		
mono-constituent substance/organic (0.44) 3,7-Diazabicyclo[3.3.1]nonane-1,...		
UVCB/organic (0.43) Fatty acids, vegetable-oil, polym...		
UVCB/organic (0.44) Disodium 5-[[4-[(2-bromo-1-ox...		
UVCB/organic (0.44) 2-(4,6-diphenyl-1,3,5-triazin-2-y...		
UVCB/organic (0.44) Oligomerisation products of ethyl...		
UVCB/organic (0.43) Alcohols, C8-10, ethoxylated		
mono-constituent substance/organic (0.43) Methyl octanoate		
mono-constituent substance/organic (0.43) ETHYLENE GLYCOL BIS(BENZENESUL...		
mono-constituent substance/organic (0.43) 17-acetoxy-1β,2β-methanopegna...		
mono-constituent substance/organic (0.43) Camphene		
mono-constituent substance/organic (0.43) 4-isobutyl-2-methylbenzaldehy...		
mono-constituent substance/organic (0.43) UK-292,679		
mono-constituent substance/organic (0.43) 2,4,6-tris(2,4,6-tribromophenoxy...		
mono-constituent substance/organic (0.43) 4-fluoroaniline		
mono-constituent substance/organic (0.42) Ethyl 4-hydroxy-2H-1,2-benzothi...		

# Substance similarity search, based on the maps (User Interface)

The screenshot displays the Ambit Substance search interface. The main search area shows a search for "pigment blue" with 0 results. A selected substance is "Disodium 4,8-diamino-1,5-dihydroxy-9,10-dioxoanthracene-2,6-disulphonate C.I. Acid Blue 45". A selected similar substance is "(1.00) Reaction products of 1,5-diamino-2,6-disulphonate with bromine Disperse Blue 056, 056:1, 056:2".

The interface includes a search bar, a "Search" button, and a "Similarity type" dropdown menu set to "structure similarity". The "Number of hits" is set to 10. The search results are displayed in a grid of cards, each showing a similarity score and a substance name. One card is highlighted in green: "(1.00) Reaction products of 1,5-diamino-2,6-disulphonate with bromine Disperse Blue 056, 056:1, 056:2".

An inset window shows a detailed view of the selected substance, "Disperse Blue 056, 056:1, 056:2". It lists various toxicity categories and their counts: 7.2.1 Acute toxicity - oral (2), 7.2.2 Acute toxicity - inhalation (1), 7.2.3 Acute toxicity - dermal (1), 7.3.1 Skin irritation / Corrosion (2), 7.3.2 Eye irritation (2), 7.4.1 Skin sensitisation (2), 7.5.1 Repeated dose toxicity - oral (1), 7.5.2 Repeated dose toxicity - inhalation (1), 7.5.3 Repeated dose toxicity - dermal (1), 7.6.1 Genetic toxicity in vitro (5), 7.8.1 Toxicity to reproduction (1), and 7.8.2 Developmental toxicity / teratogenicity (1). A blue arrow points from the "7.3.2 Eye irritation (2)" entry in the inset to the "structure similarity" dropdown in the main interface.

Query

Choose property

- P-CHEM Vapour pressure VAPOUR\_PRESSURE Pa (23356)
- P-CHEM Partition coefficient LOG\_POW (34622)
- P-CHEM Melting point/freezing point MELTING\_POINT °C (30426)
- P-CHEM Water solubility WATER\_SOLUBILITY mg/L (20839)
- P-CHEM Boiling point BOILING\_POINT °C (34500)

Source

Choose quantile

- 70
- 80
- 90
- 100

Show 10 entries

Search: (Q090\_s:PC\_WATER\_SOL\_SECTION\_f9a

Description	Substance type	Q010	Q020	Q030	Q040	Q050	Q060	Q070	Q080	Q090	Q100	Find similar by
[3-(2,3-epoxypropoxy)propyl]diethoxymethylsilane KBE-402	mono-constituent substance											property Qstructure
N-[3-(trimethoxysilyl)propyl]aniline	mono-constituent substance											property Qstructure
2-chloropyridine Halogenated pyridine	mono-constituent substance											property Qstructure
2-(3,4-epoxycyclohexyl)ethyltrimethoxysilane	mono-constituent substance											property Qstructure
Reaction product of alkylaryl sulphonic acid and alkanolamine	UVCB											property Qstructure
N,N-diethyl-3-oxobutyramide	mono-constituent substance											property Qstructure
Thiosemicarbazide	mono-constituent substance											property Qstructure
[3-(2,3-epoxypropoxy)propyl]triethoxysilane	mono-constituent substance											property Qstructure



# AMBIT 3 LRI Read Across Tool

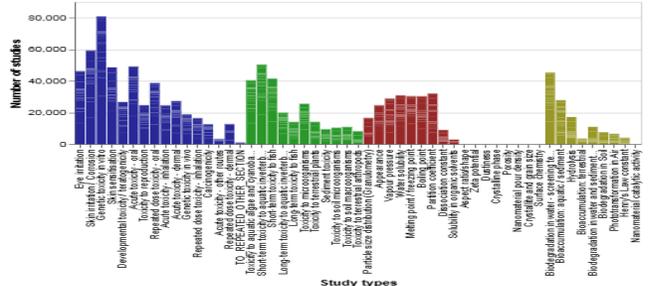
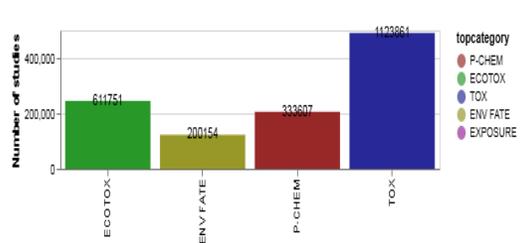
The screenshot shows the LRI Read Across tool interface with three main sections: Search, Read across, and Predict. The Search section includes a search bar and filters for structures, substances, and studies. The Read across section offers options to create or explore assessments. The Predict section allows for applying predictive models. A sidebar on the left lists data sources like REACH Study Results and OpenFoodTox, and a legal notice section.

The screenshot shows the GitHub repository for 'notebooks-ambit' by 'ideaconsult'. It displays the repository name, star count, and a list of notebooks including 'toxtree models', 'vega models', and 'ambit-knime'. The 'ambit-knime' notebook is highlighted, showing its README content.

The screenshot shows the documentation for 'jToxKit', a JavaScript library for chem-informatics UI and data management. It includes a background section and a list of features such as multi-layered architecture, LUGO-like principles, and a modular design.

## REST API & libraries

A blue box containing the command `pip install pynanomapper 1.0.4` with a document icon.



<https://ambitlri-dev.ideaconsult.net>  
[support@ideaconsult.net](mailto:support@ideaconsult.net)

# AMBIT 3 LRI Read Across Tool

<https://ambitlri-dev.ideaconsult.net/>

The screenshot shows the AMBIT 3 LRI Read Across tool interface. It has a green header with the logo and title. The main content area is divided into several sections: a search bar, a 'Read across assessments' section with a 'Read across' button, a 'Chemical structure properties' section with a 'Predict' button, and a 'Data sources' section listing various data sources. A 'Legal notice' section is also present at the bottom left.

## Frontend updates

- New responsive design
- Upgrade of existing user interface
  - Read Across, Structure search
- New user interface
  - Dashboard, substance similarity

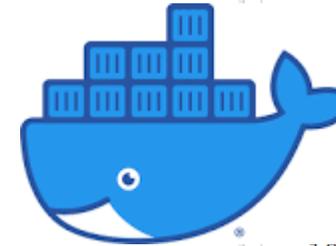
## New functionalities

- IUCLID6.6 support
- VEGA integration through REST API
- Data maps similarity search

## Backend updates:

- New data integration approach through search engine (metadata, data and multi property similarity)
- Modular (microservices)

The screenshot shows a GitHub repository page for 'ideaconsult / ambit-docker'. The repository structure is visible, including files like 'data\_import', 'README.md', 'ambit-config.env', and 'docker-compose.yml'. The README.md file is open, showing the title 'The Cefic-LRI cheminformatics tool' and instructions for getting started with Docker Compose.



<https://github.com/ideaconsult/ambit-docker/>

Distribution using modern containerized approach



# Thank you!

Questions?

