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

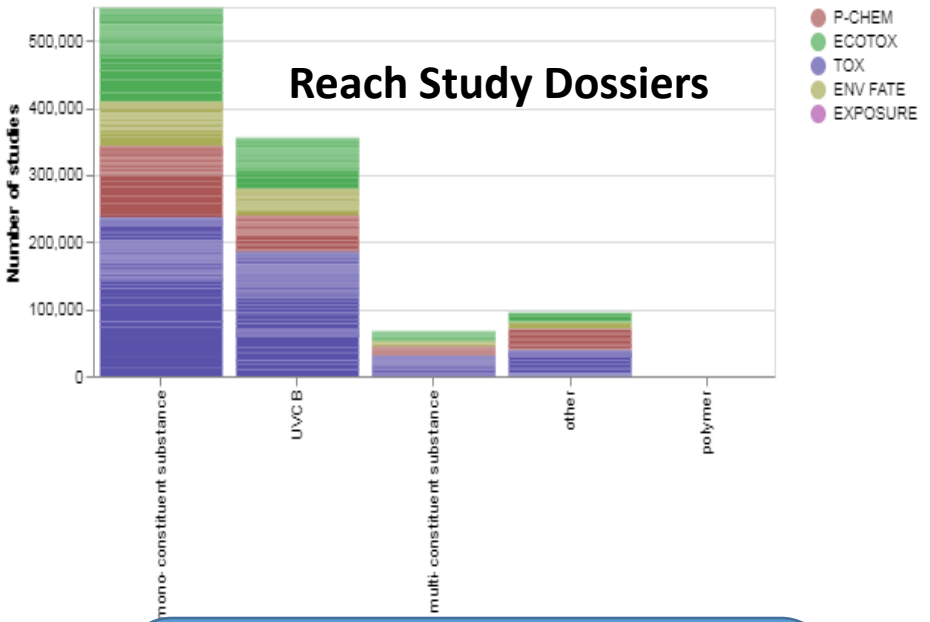
Dr. Nina Jeliazkova

Ideaconsult Ltd. , Sofia, BG

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

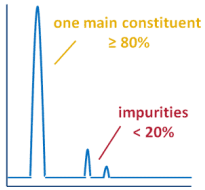
What's in a substance ?

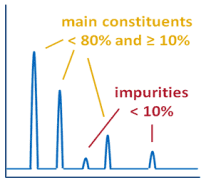
IUPAC INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

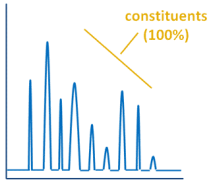
IUPAC definition:
a substance is a matter of constant composition, best characterised by the entities (molecules, formula units, atoms) it is composed of.

ECHA

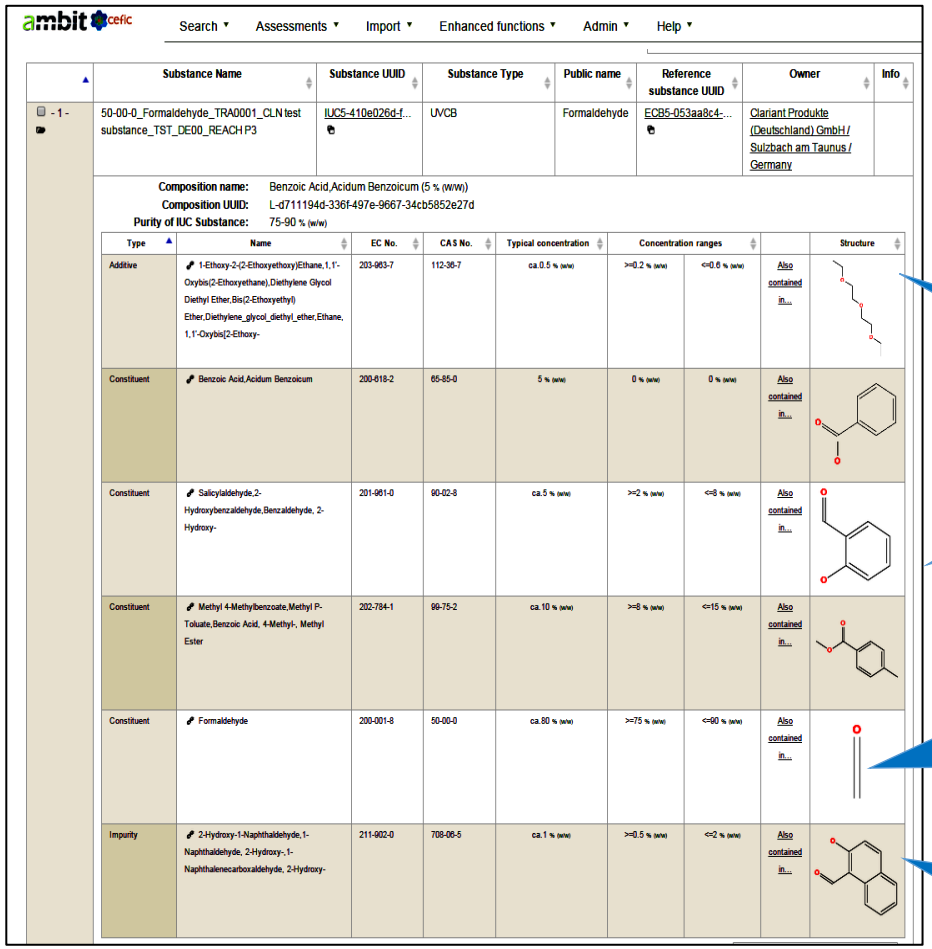
ECHA Substance Identification
A substance is a chemical element and its compounds in the natural state or the result of a manufacturing process.







ambit



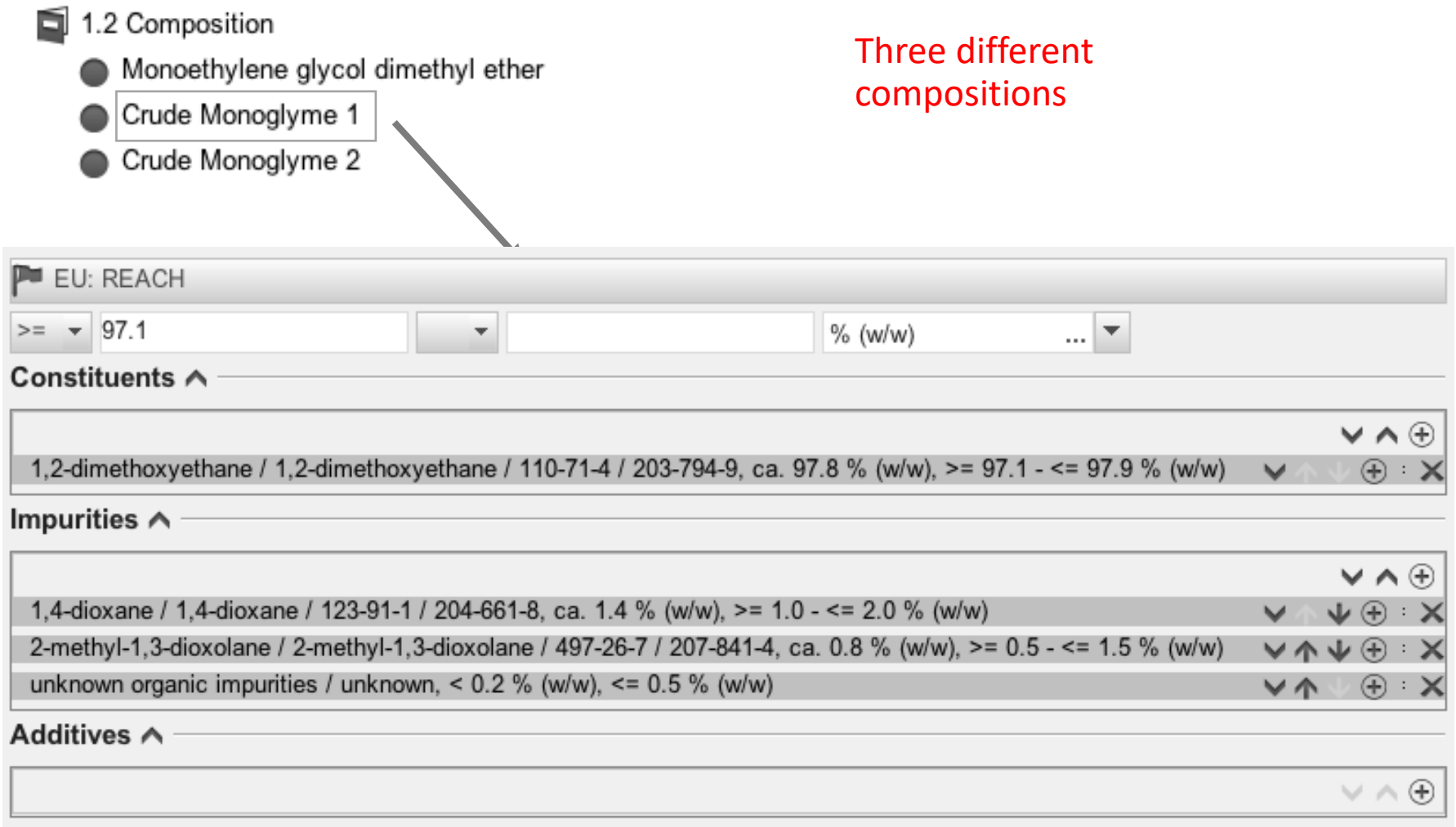
Additive

Constituent

Main constituent

Impurity

REACH substance definition as implemented in IUCLID
Example: mono-constituent



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.

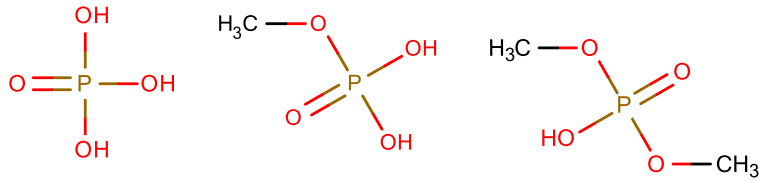
> ***Main constituent***: A constituent, not being an additive or impurity, in a substance that makes up a significant part of that substance. Contributes to the naming of the substance. Concentration of the main constituent(s) = purity of the substance.

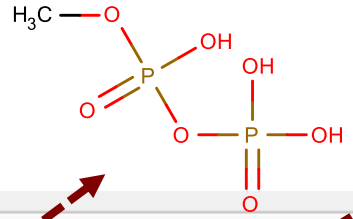
Additive: A substance that has been intentionally added to stabilise the substance. Contributes to the substance composition.

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





H_2O

>= 80.0 <= 100.0 % (w/w) ...

Constituents

orthophosphoric Acid / phosphoric acid / 7664-38-2 / 231-633-2, , >= 20.0 - <= 55.0 % (w/w)

methyl dihydrogen phosphate / methyl dihydrogen phosphate / 812-00-0 / 212-379-1, , >= 25.0 - <= 75.0 % (w/w)

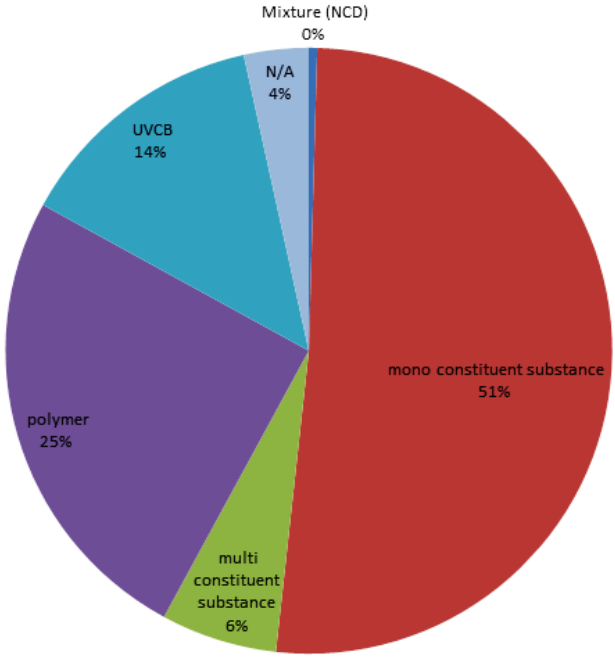
dimethyl hydrogen phosphate / dimethyl hydrogen phosphate / 813-78-5 / 212-389-6, , >= 0.0 - <= 10.0 % (w/w)

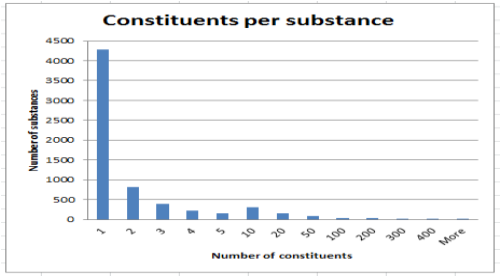
Impurities

methanol / methanol / 67-56-1 / 200-659-6, , >= 0.0 - <= 2.0 % (w/w)

pyro phosphoric acid, methyl esters / pyro phosphoric acid, methyl esters, , >= 0.0 - <= 3.0 % (w/w)

water / dihydrogen oxide / 7732-18-5 / 231-791-2, , <= 15.0 % (w/w)





non-confidential REACH data supplied by ECHA

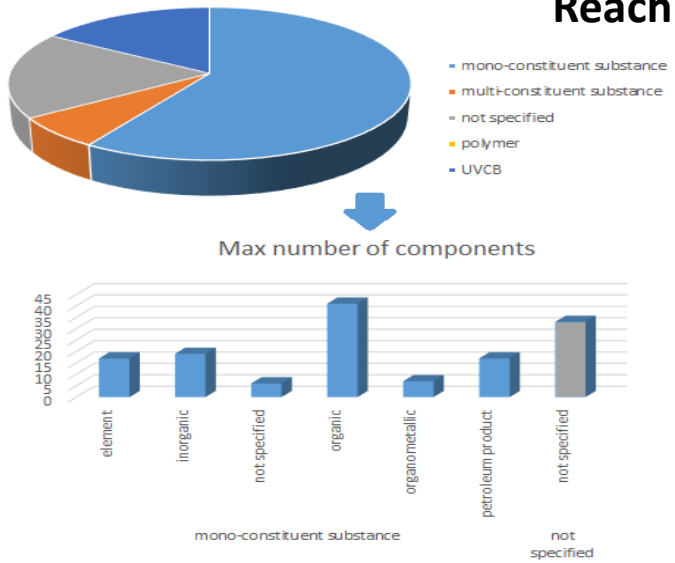
http://ambitlri.ideaconsult.net

Chemical substance database (background)

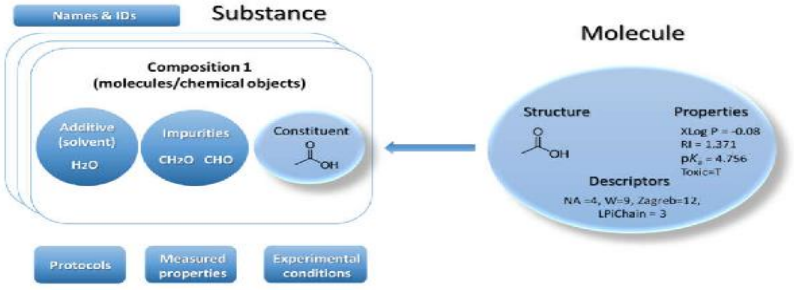
ambit

- 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

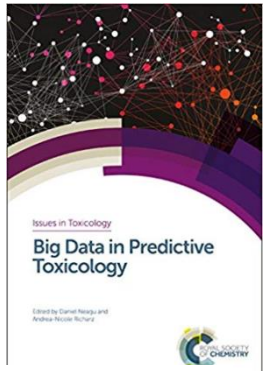
Reach Study Dossiers



it takes more than a single structure to describe a material



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



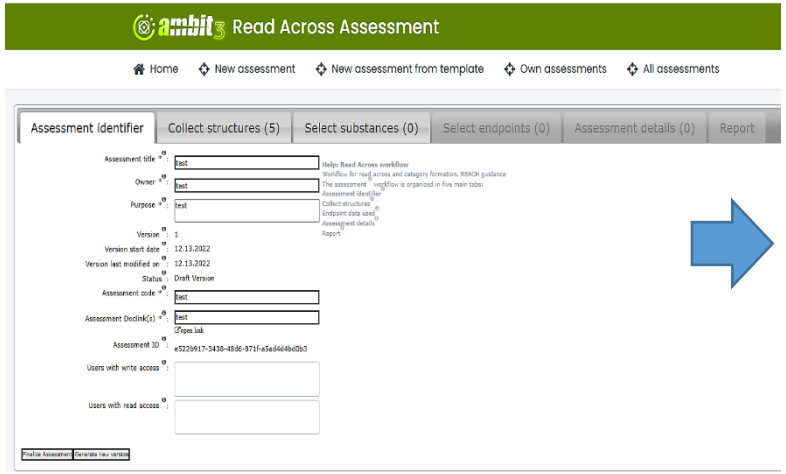
CHAPTER 3

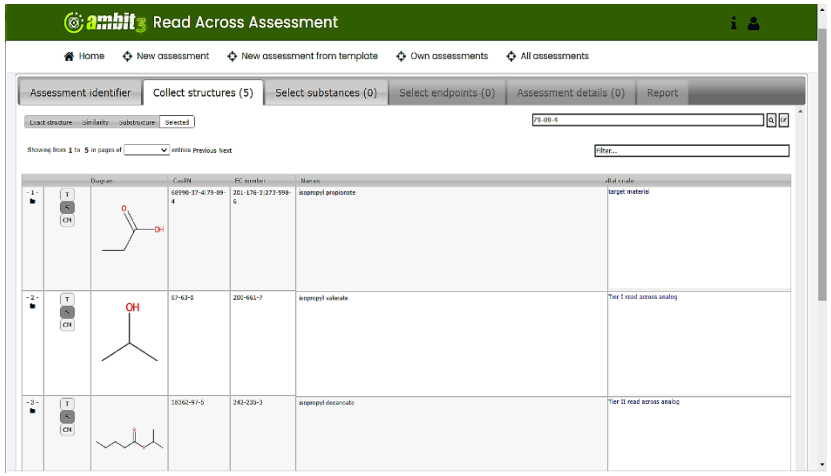
Chemoinformatics Representation of Chemical Structures – A Milestone for Successful Big Data Modelling in Predictive Toxicology

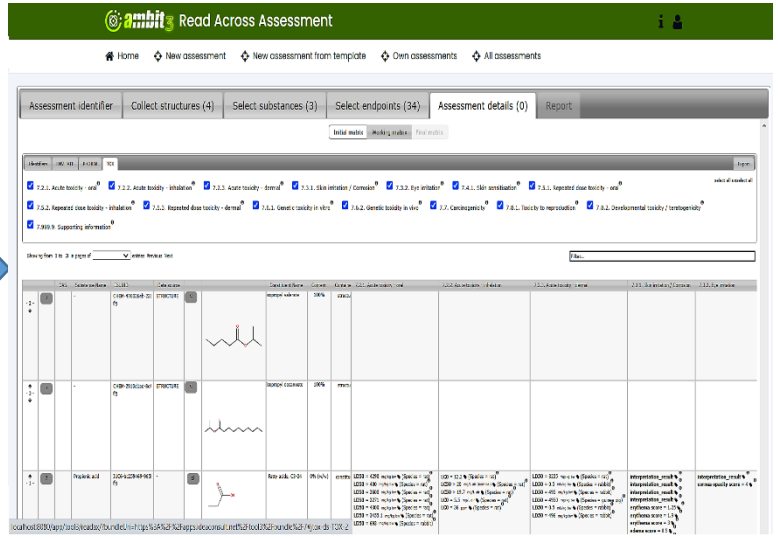
Nikolay Kochev, Nina Jeliazkova and Ivanka Tsakovska

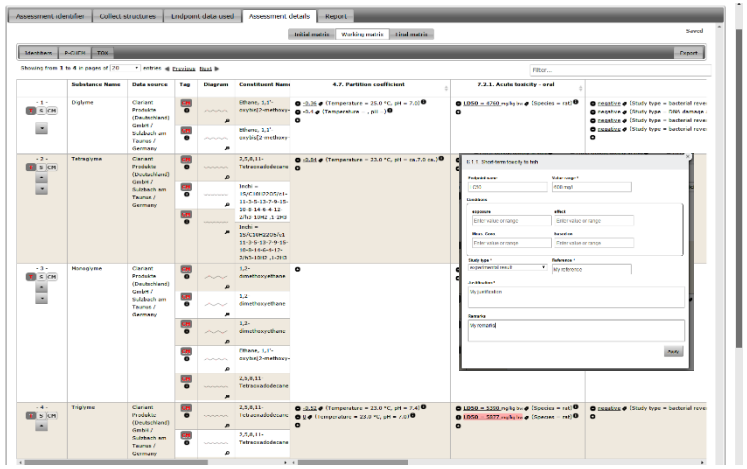
LRI Read Across Tool

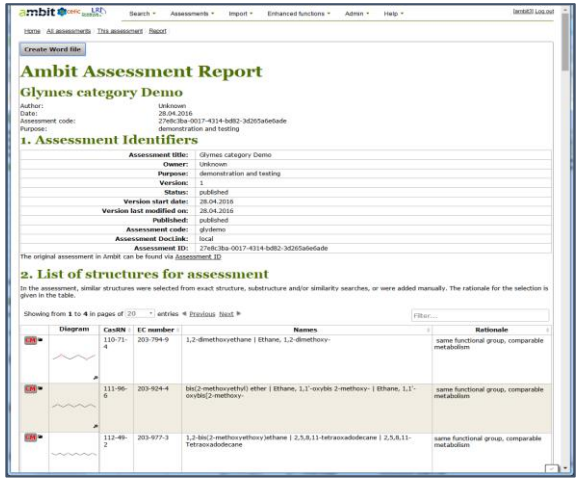
https://ambitlri.ideaconsult.net

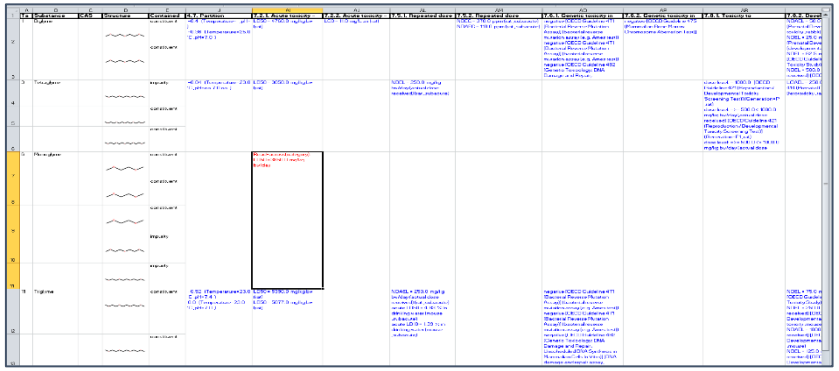




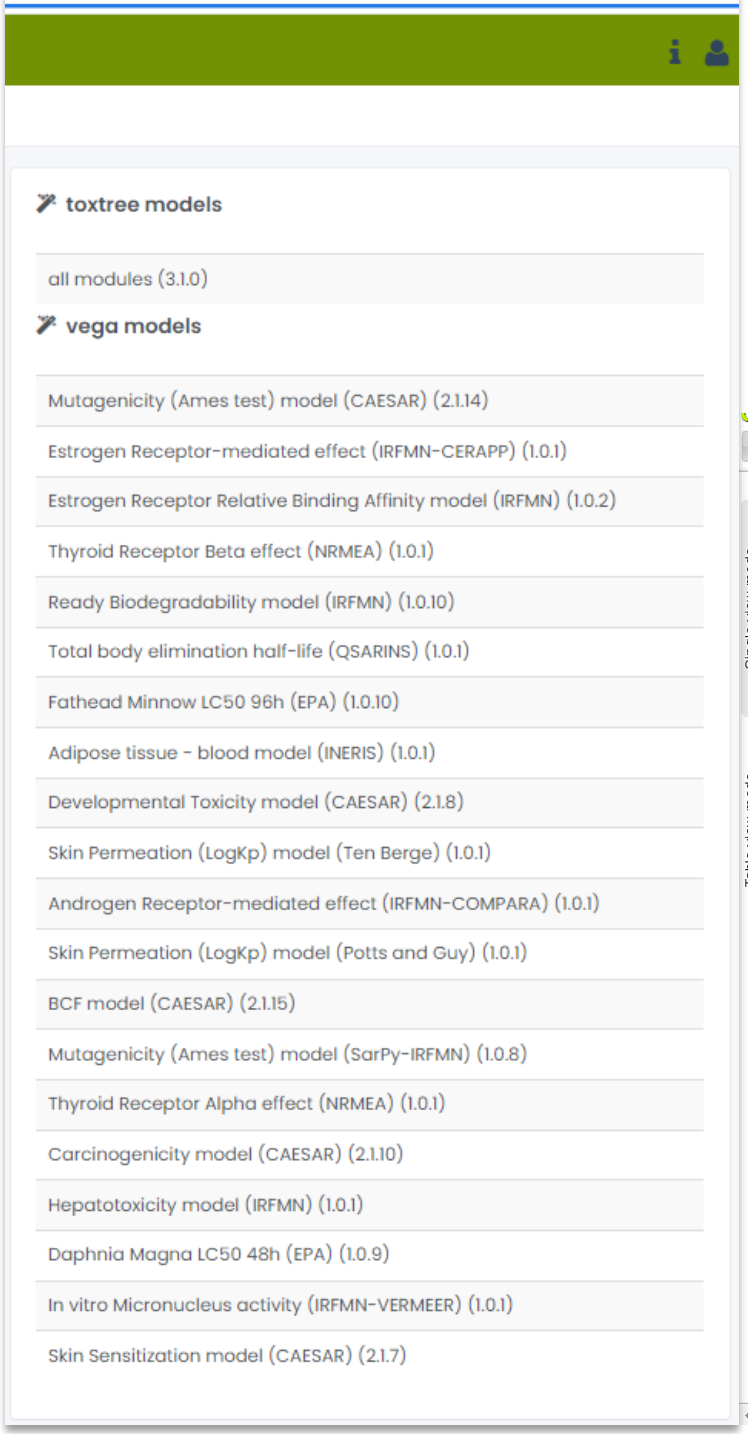


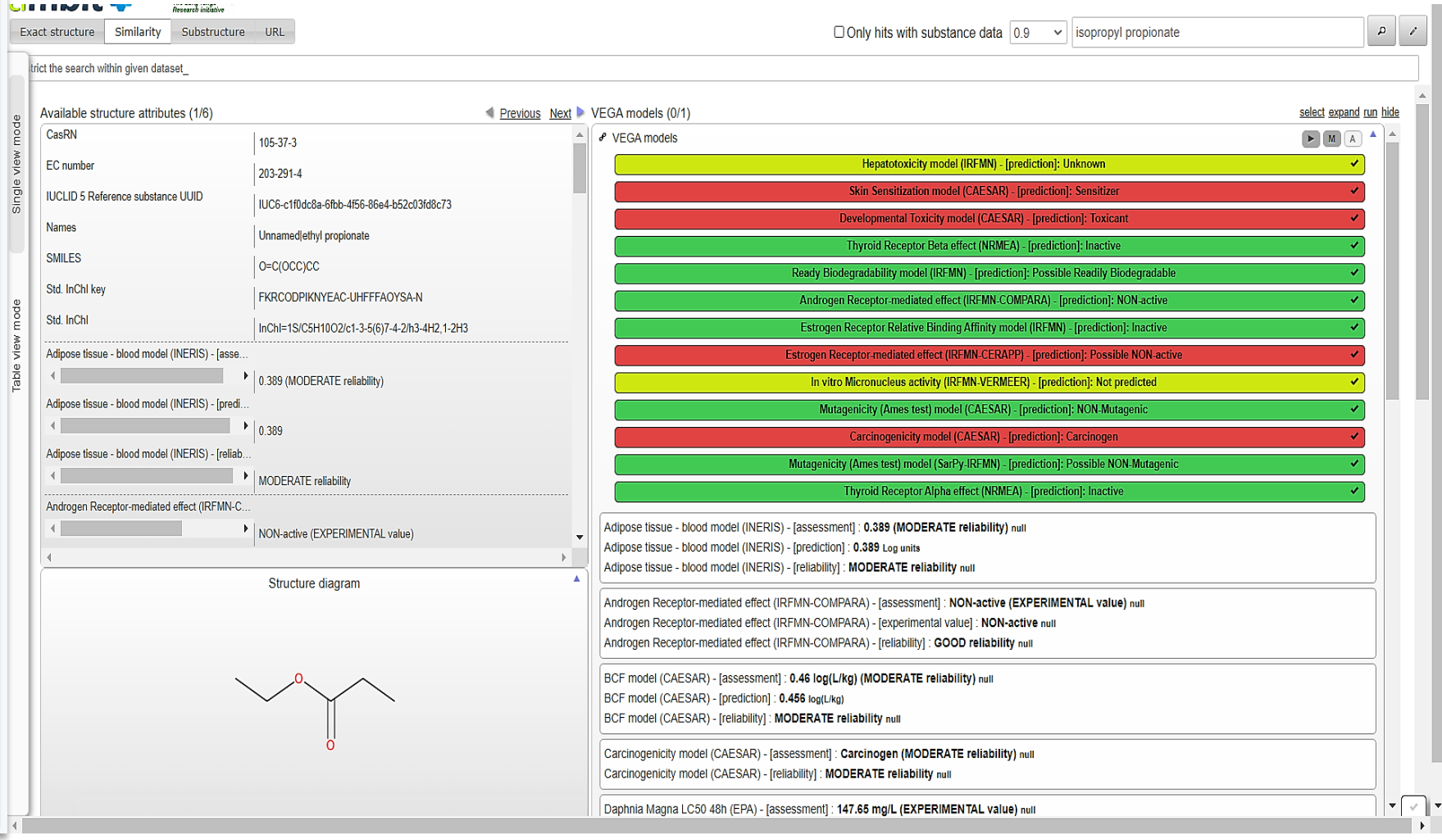






Toxtree (toxtree.sf.net)
Vega integration (collab with IRFMN)





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

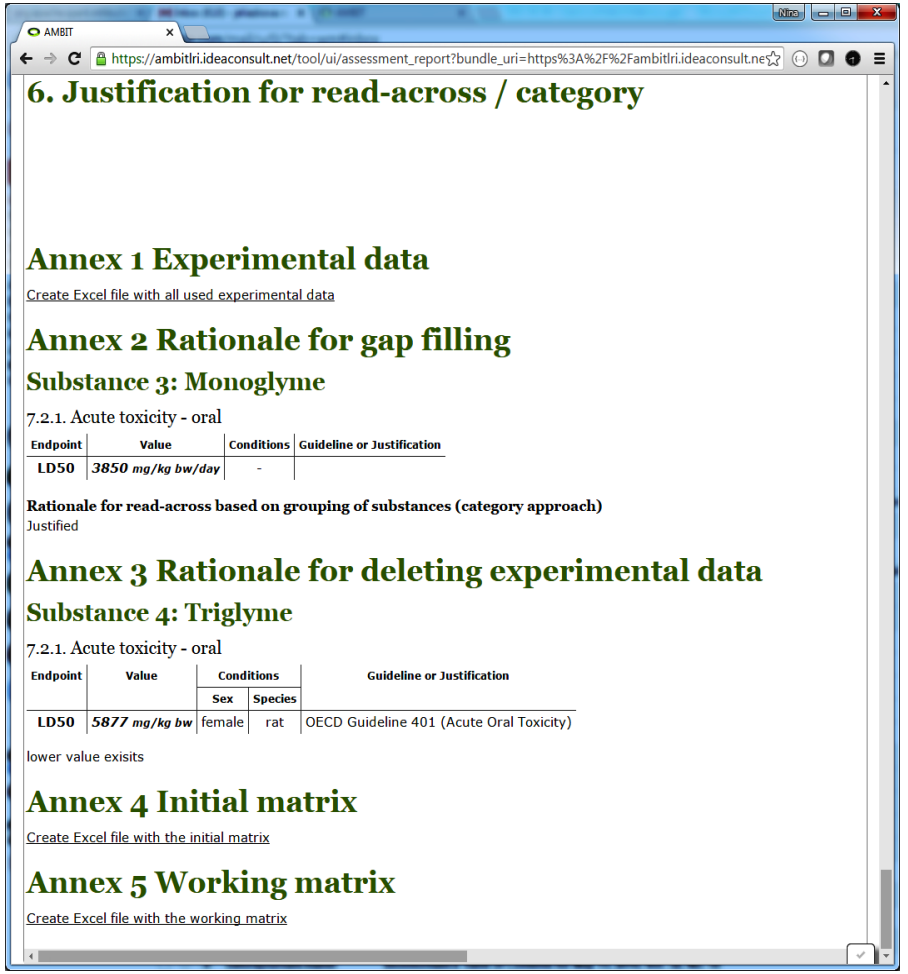
5. Assessment data matrix

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

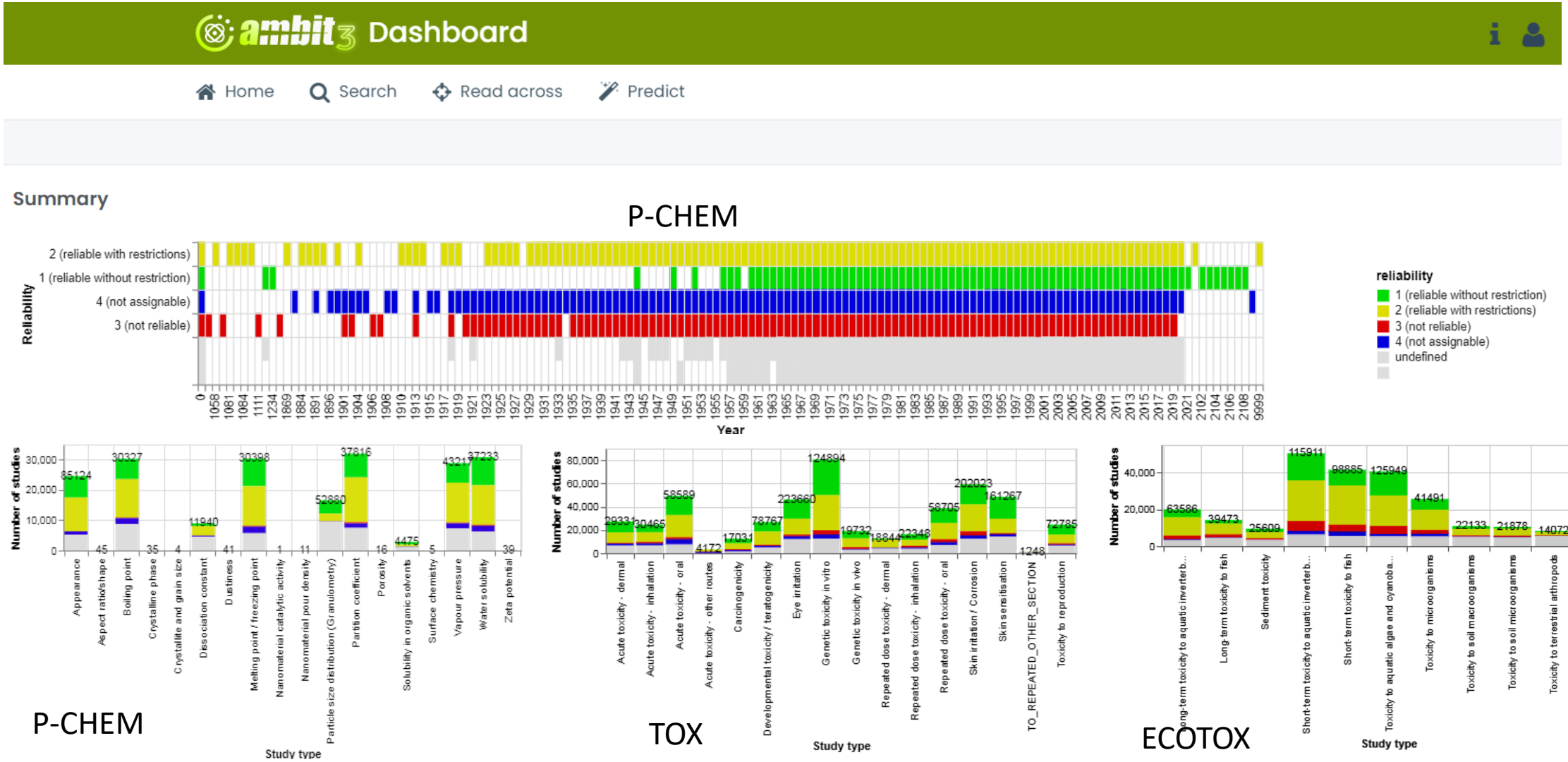
For detailed data or rationale for waiving and read-across, click hyperlinks in the table. These data or rationales can also be found in the annex of the report.

Tag	CM CM	CM CM CM	
Substance name	**Diglyme**	**Tetraglyme**	
CAS No.			
4.7. Partition coefficient	-0.36 (Temperature = 25.0 °C, pH = 7.0) -0.4 (Temperature = , pH =)	-0.84 (Temperature = 23.0 °C, pH = ca.7.0 ca.)	
7.2.1. Acute toxicity - oral	LD50 = 4760 mg/kg bw (Species = rat)	LD50 = 3850 mg/kg bw (Species = rat)	LD5
7.2.2. Acute toxicity - inhalation	LC0 = 11 mg/L air (Species = rat)		
7.2.3. Acute toxicity - dermal			
7.3.1. Skin irritation / Corrosion	erythema score = 0.89 edema score = 0.5	erythema score = 0.8 edema score = 0 edema score = 0 erythema score = 1.3 erythema score = 1 edema score = 0 edema score = 0 erythema score = 0	
7.3.2. Eye irritation	Maximum mean total score (MMTS) = 13 cornea score = 0 iris score = 0 conjunctivae score = 0.78 chemosis score = 0.28	chemosis score = 0 chemosis score = 0 iris score = 0 cornea score = 0 cornea score = 0 iris score = 0 chemosis score = 0 conjunctivae score = 0.3 cornea score = 0 iris score = 0 conjunctivae score = 0 chemosis score = 0 cornea score = 0 conjunctivae score = 0.1 iris score = 0 conjunctivae score = 0	
7.5.1. Repeated dose toxicity - oral		NOEL = 250 mg/kg bw/day (actual dose received) (Species = rat, Test type = subacute)	
7.5.2. Repeated dose toxicity - inhalation	NOEC = 370 ppm (Species = rat, Test type = subacute) NOAEC = 110 ppm (Species = rat, Test type = subacute)		
7.6.1. Genetic toxicity in vitro	negative (Study type = bacterial reverse mutation assay (e.g. Ames test), OECD		

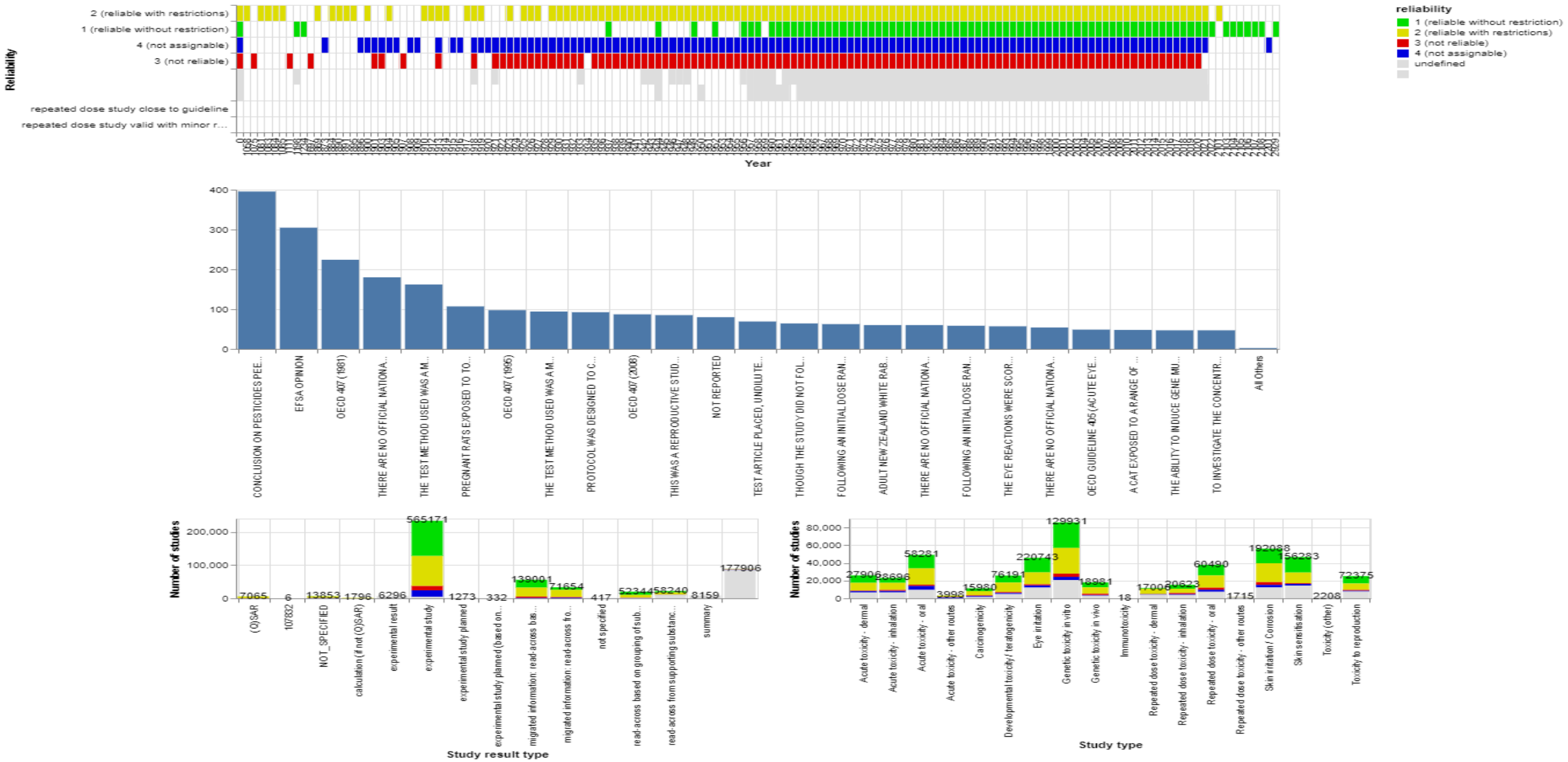
- The report describes the read across workflow in 6 sections and 5 annexes



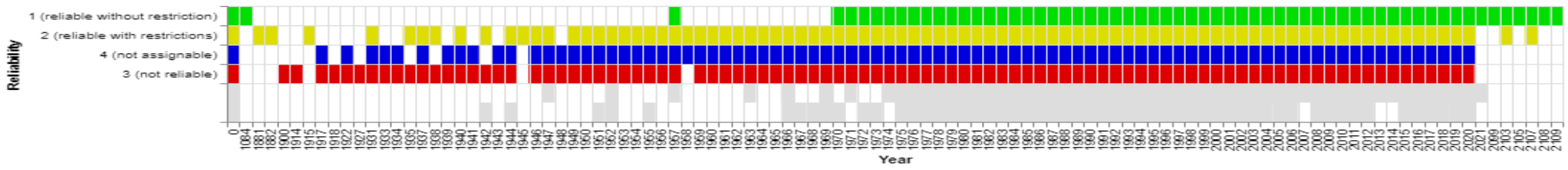
Dashboards : Data quality (REACH dossiers)

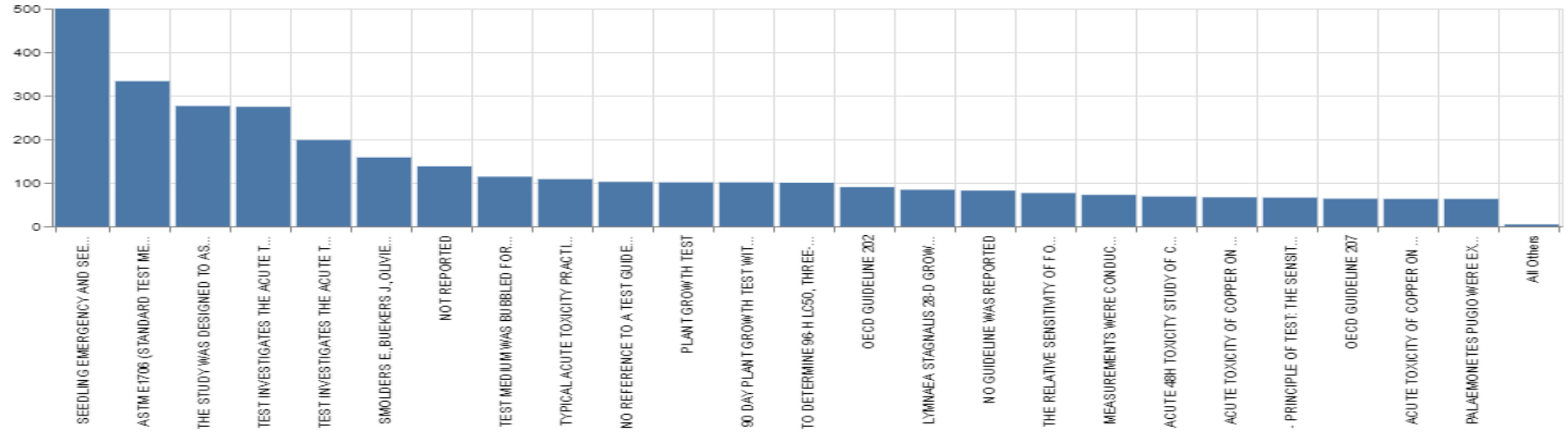


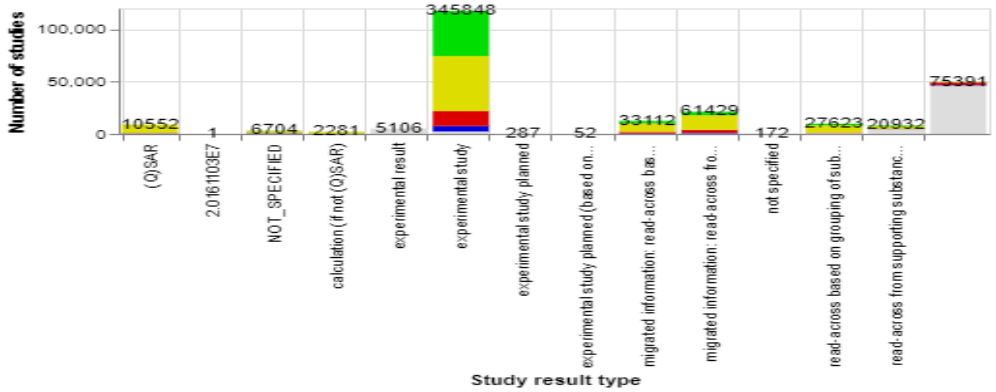
Dashboards : TOX Data quality (REACH dossiers)

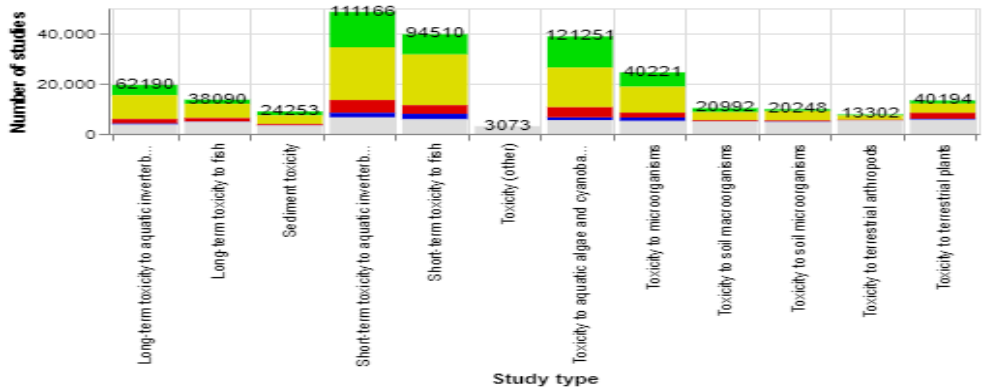


Dashboards : ECOTOX Data quality (REACH dossiers)

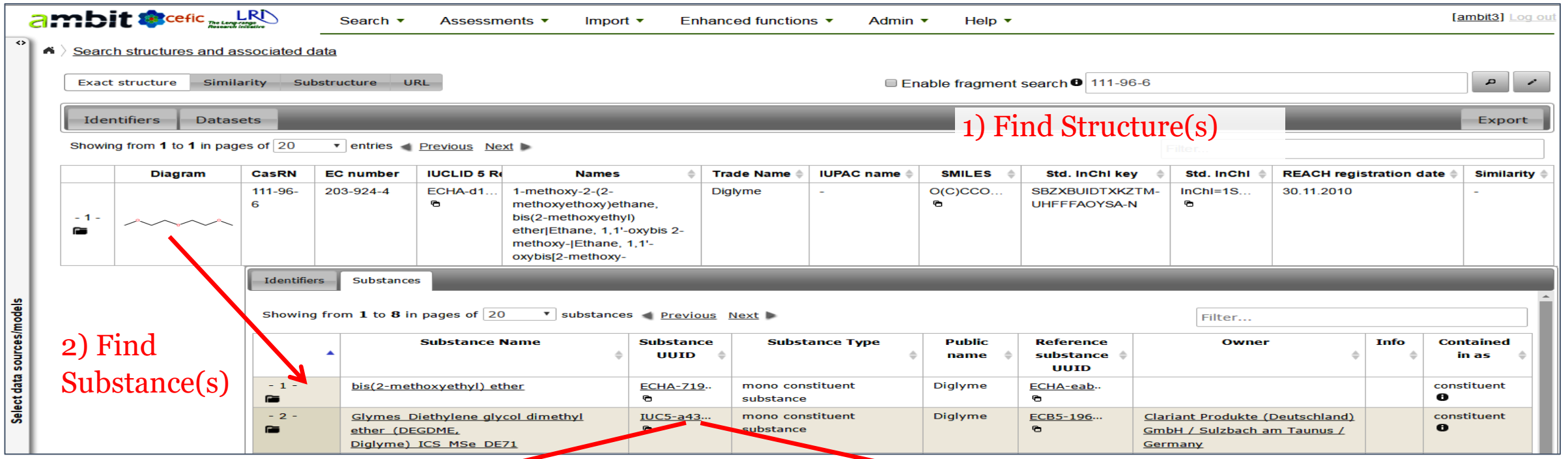






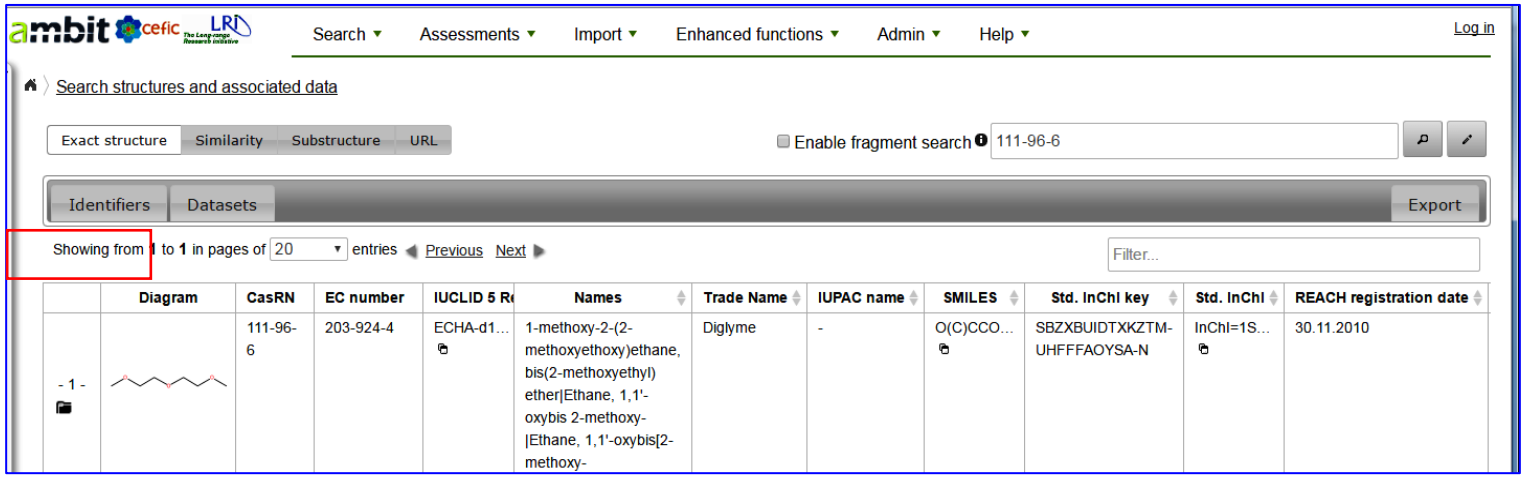


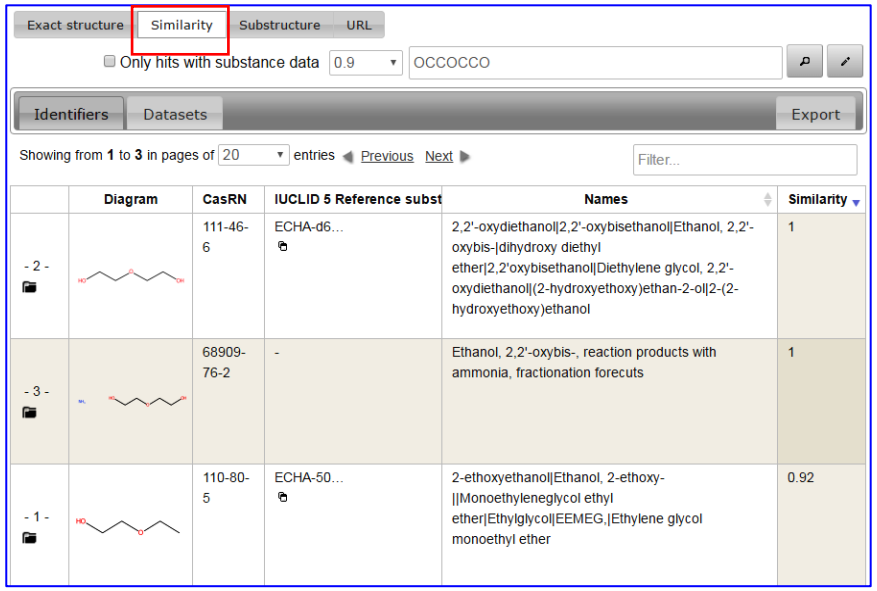
AMBIT Search for Structures & Endpoint data

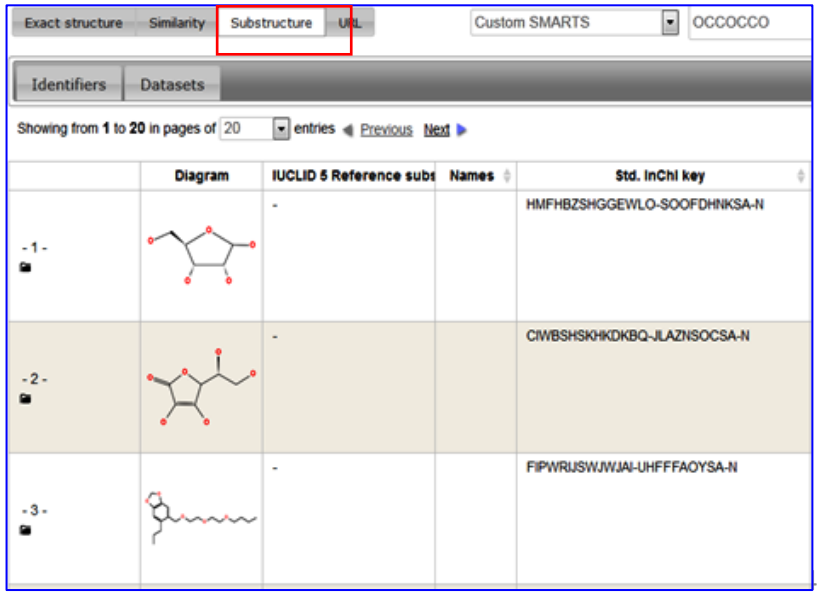


AMBIT Search for Structures

Exact, Similarity & Substructure

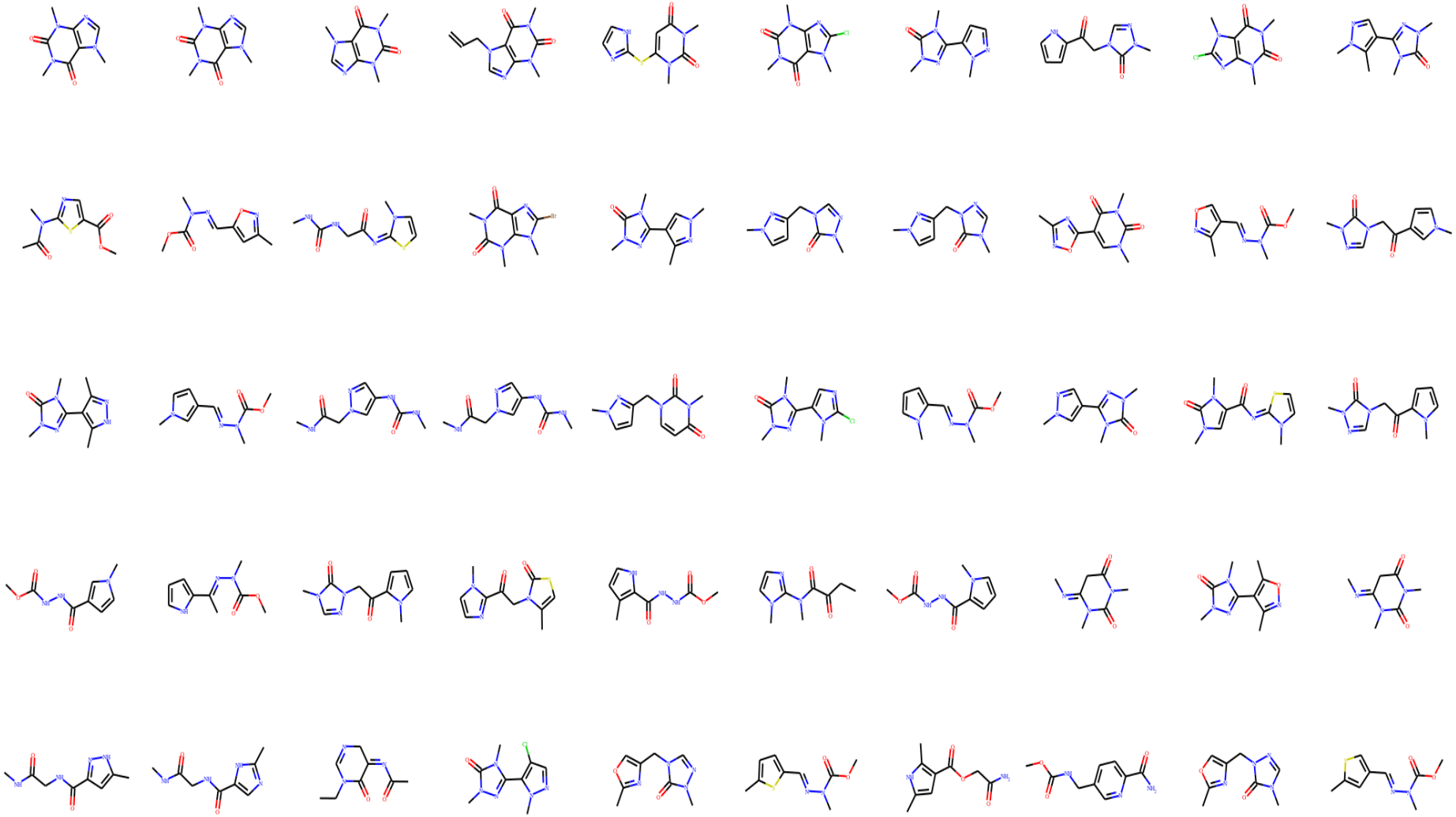


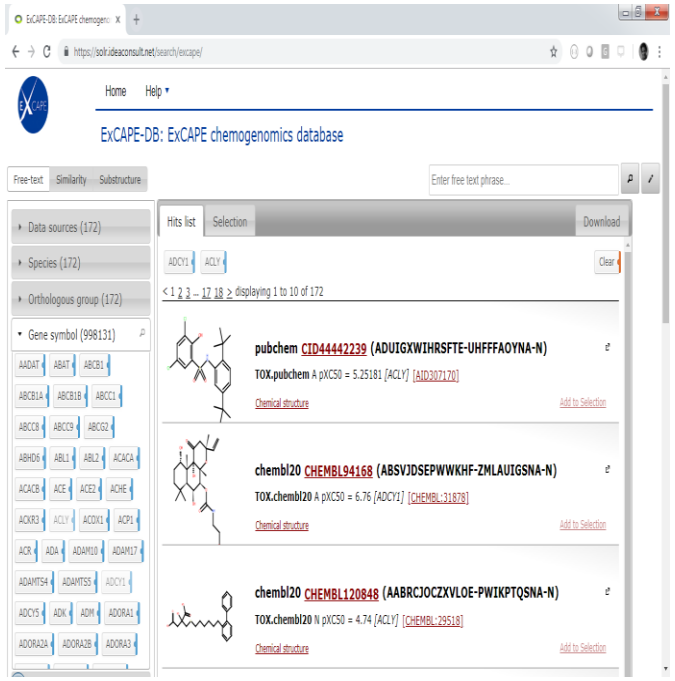




Structure similarity search (>300 mln structures)

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





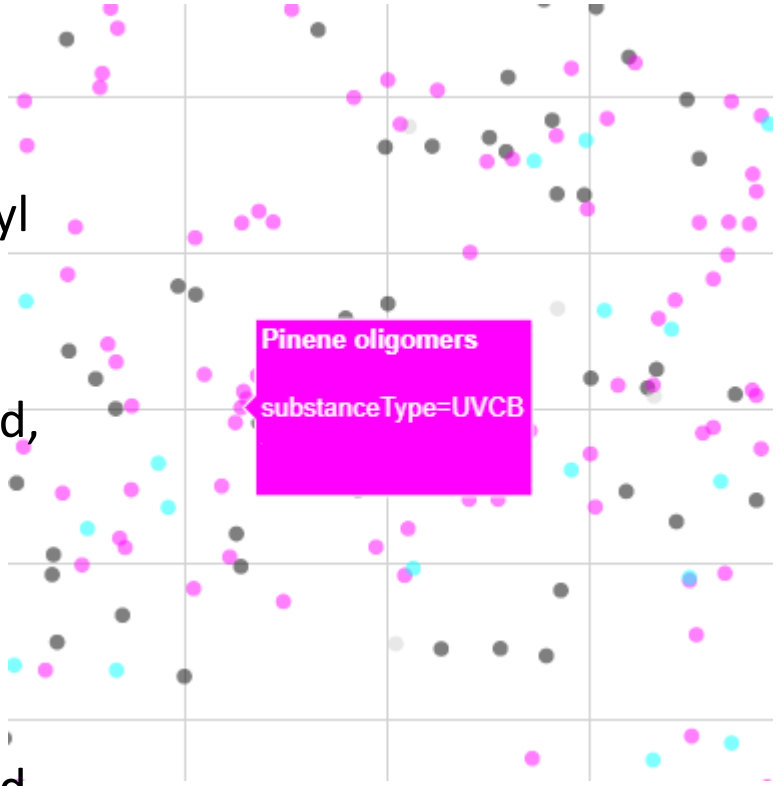
Industry-scale application and evaluation of deep learning for drug target prediction

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

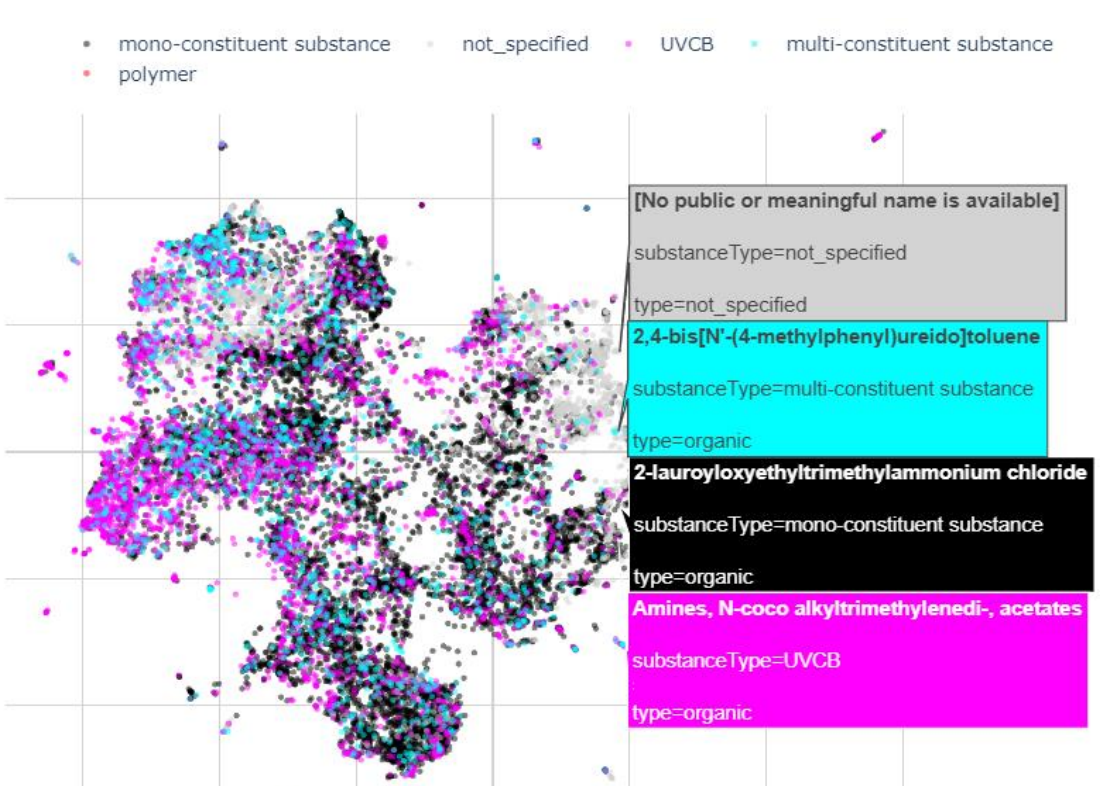
Journal of Cheminformatics **12**, Article number: 26 (2020) | Cite this article

AMBIT3 : Substance similarity query (example)

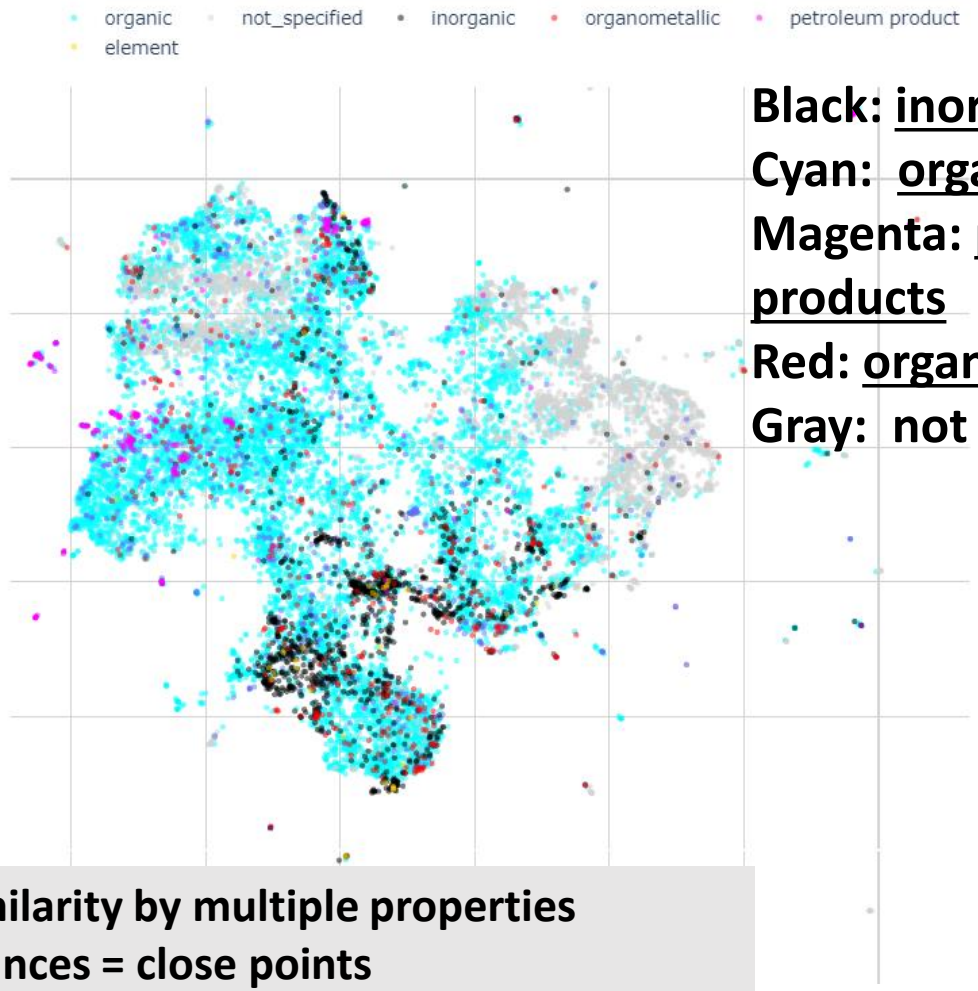
- **Query**: **Pinene oligomers (UVCB)**
- **Hits**:
  - Benzenesulfonic acid, 4-mono-C20-24 (even numbered)-alkyl derivs., magnesium salts Benzenesulfonic acid, mono-C20-24-alkyl derivs., magnesium salts
  - Tetradecyl methacrylate 2-Propenoic acid, 2-methyl-, tetradecyl ester
  - Decanoic acid, mixed esters with dipentaerythritol, heptanoic acid, isononanoic acid, and octanoic acid
  - Benzenesulfonic acid, 4-(branched alkyl derivs.) and benzenesulfonic acid, 4-(linear alkyl dervis.), 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)



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

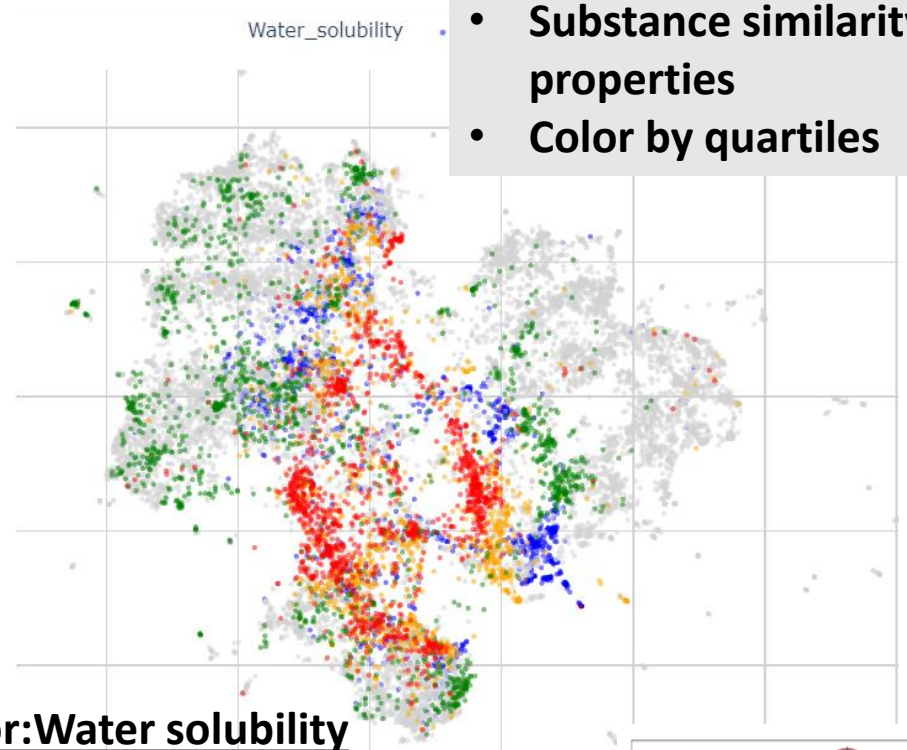


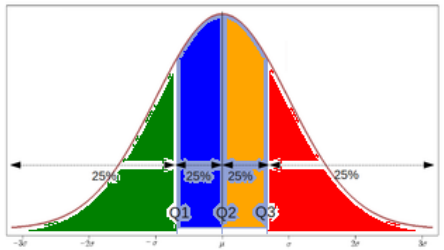
Black: inorganic
Cyan: organic
Magenta: petroleum products
Red: organometalic
Gray: not specified

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

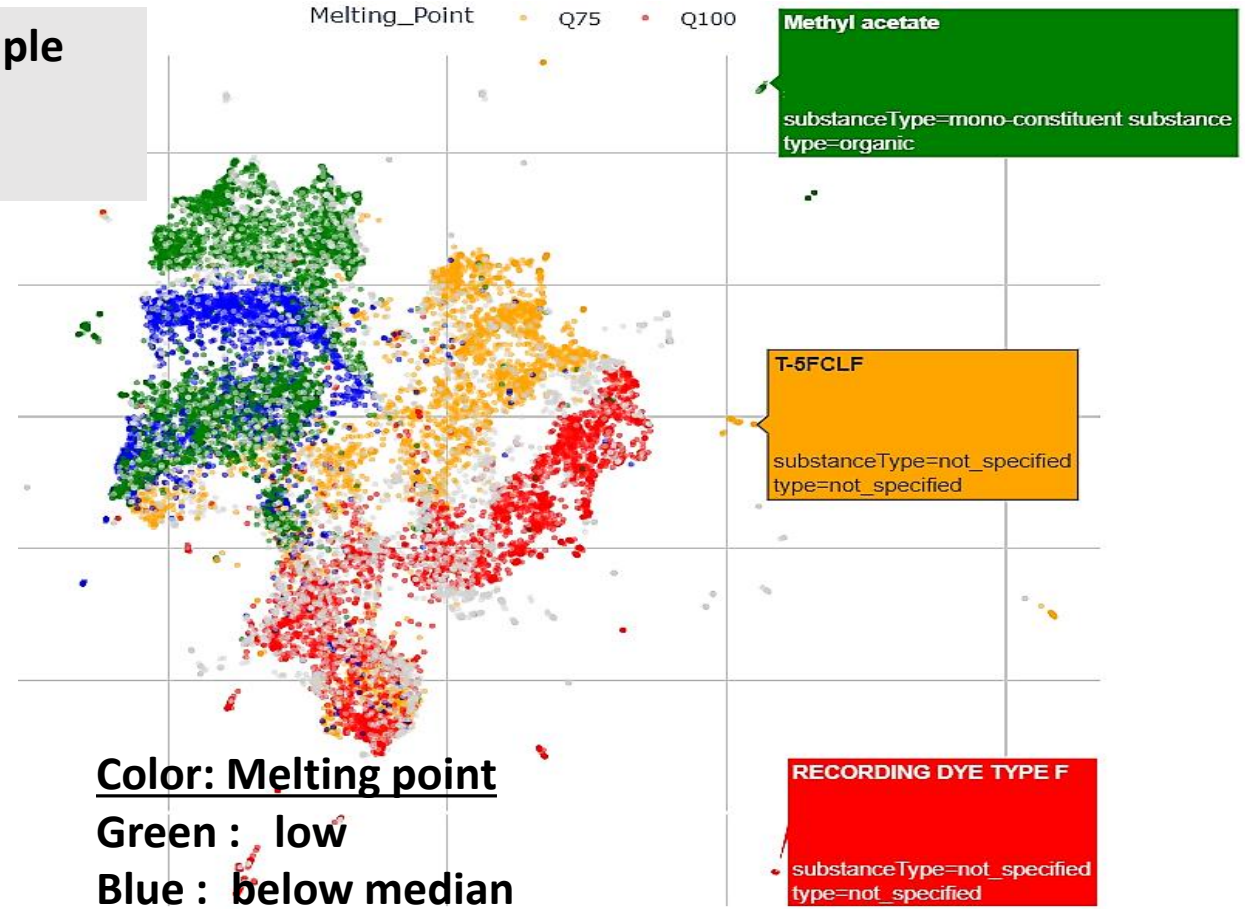
AMBIT 3 Data maps (REACH dossiers)

- **Substance similarity by multiple properties**
- **Color by quartiles**





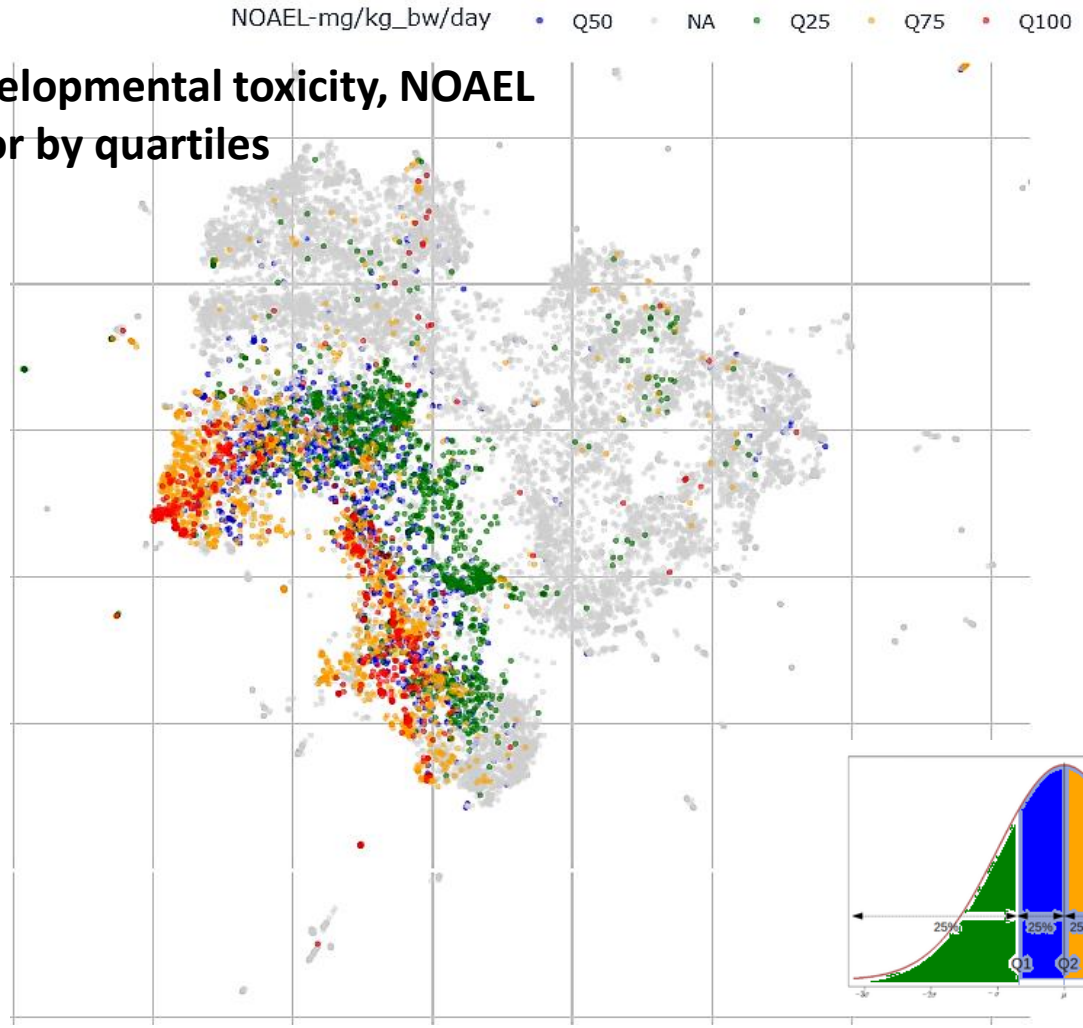
Color:Water solubility
Green : low
Blue : below median
Orange : above median
Red : highest water solubility



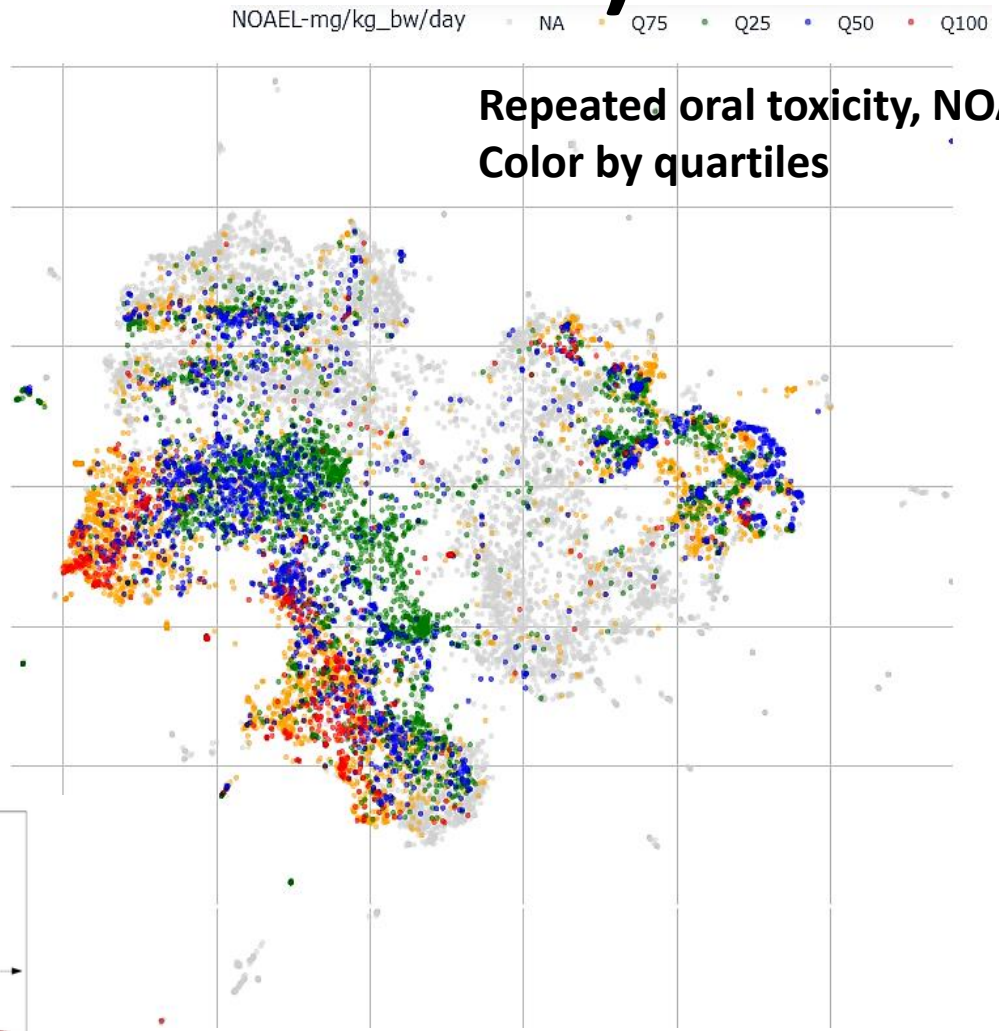
Color: Melting point
Green : low
Blue : below median
Orange : above median
Red : highest melting point

AMBIT 3 Data maps (REACH dossiers)

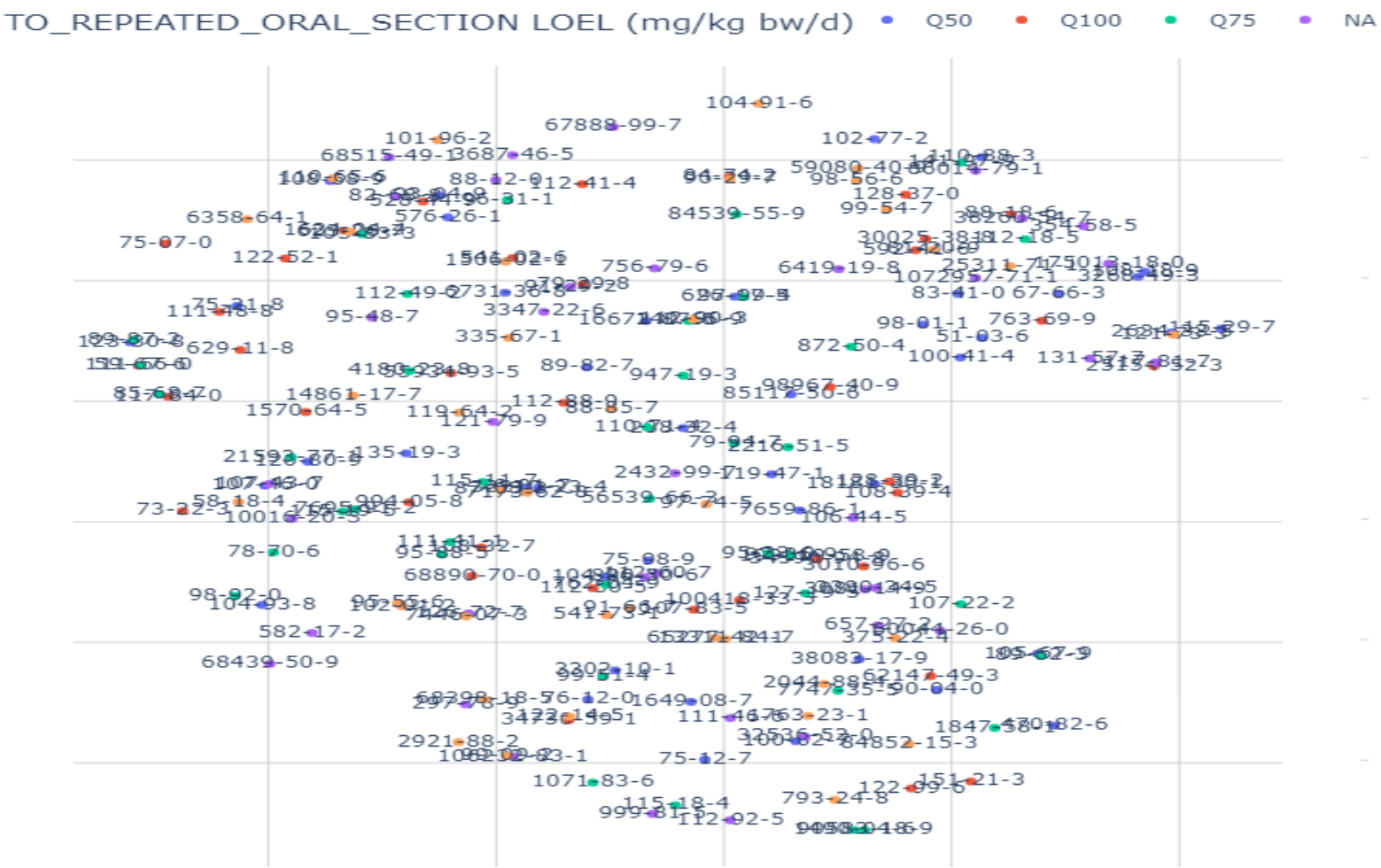
Developmental toxicity, NOAEL
Color by quartiles



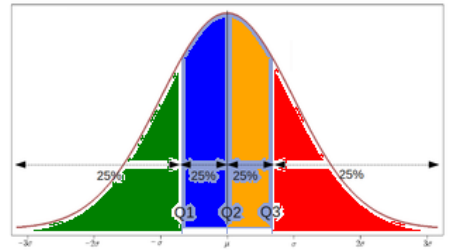
Repeated oral toxicity, NOAEL
Color by quartiles



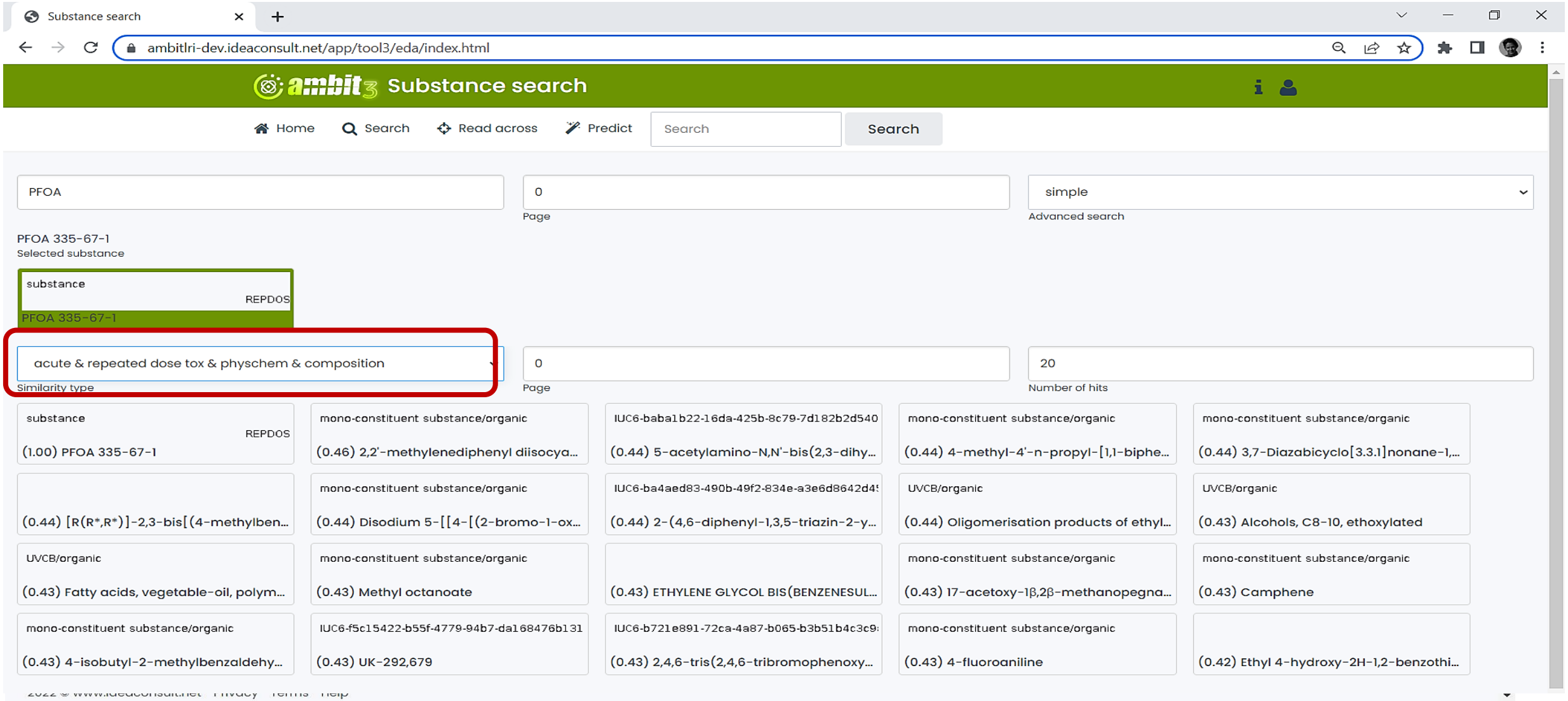
AMBIT 3 Data maps (REPDOSE , Fraunhofer ITEM)



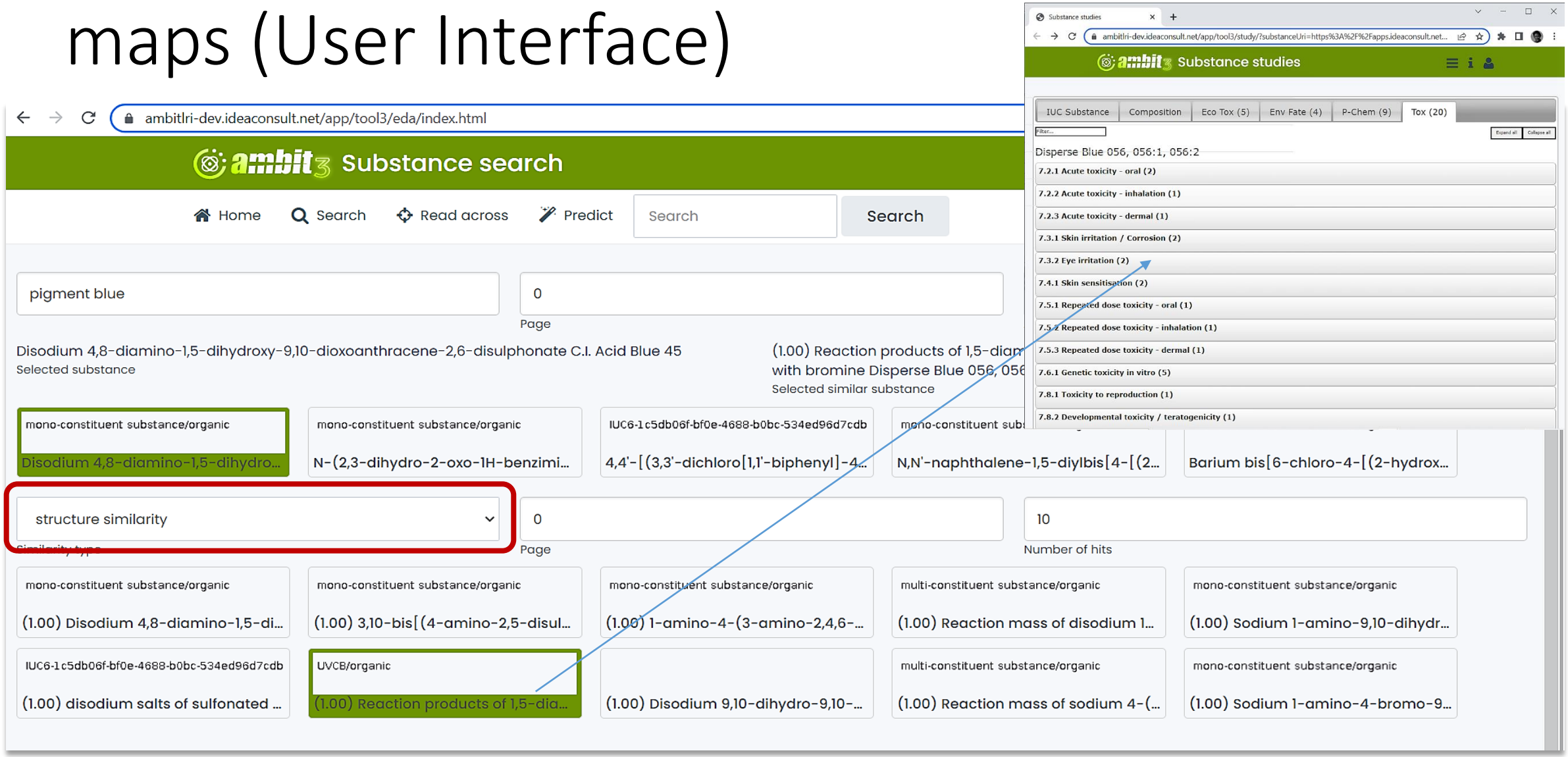
Repeated oral toxicity, LOEL
Color by quartiles



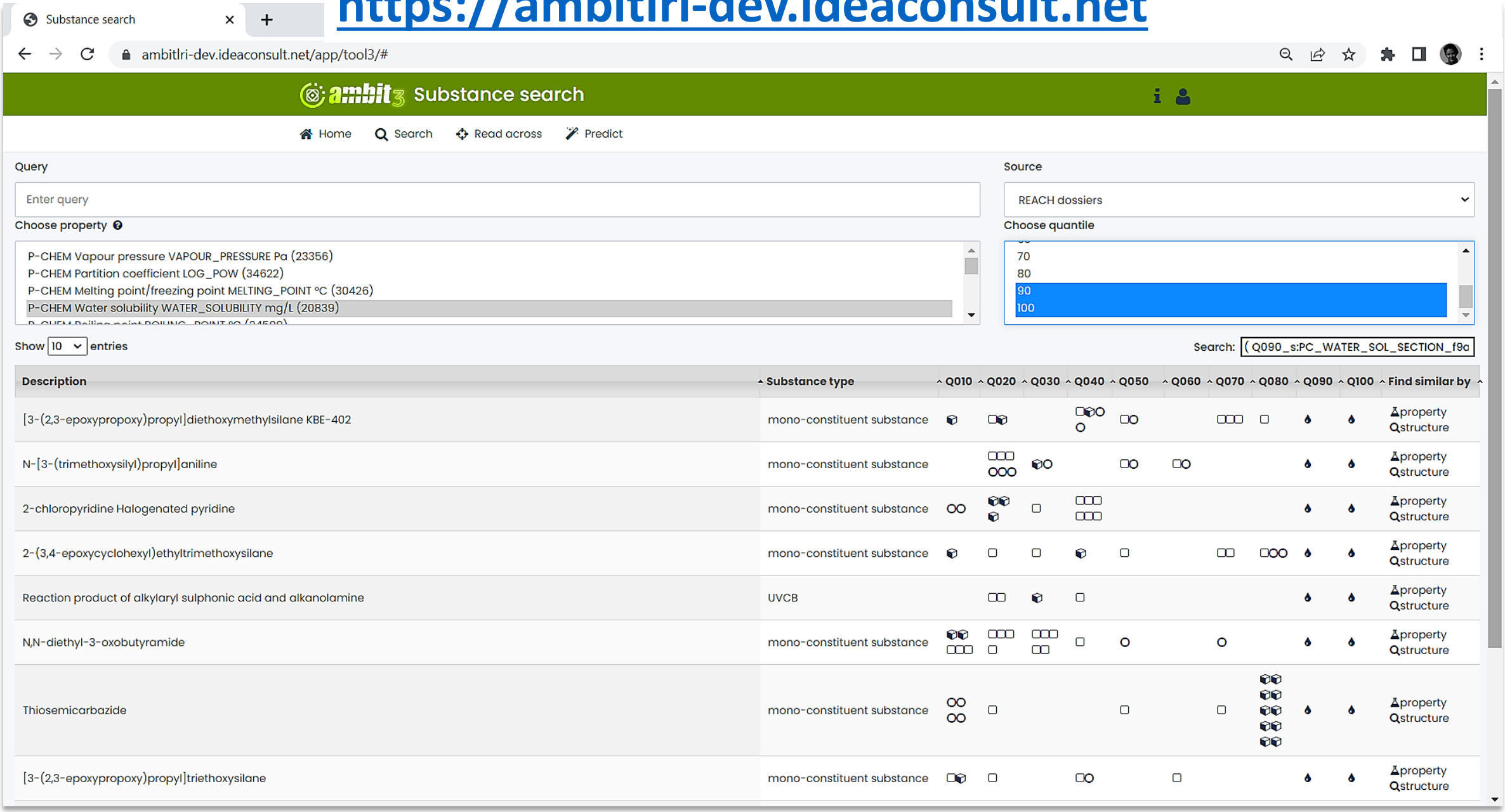
Substance similarity search, based on the maps (User Interface)



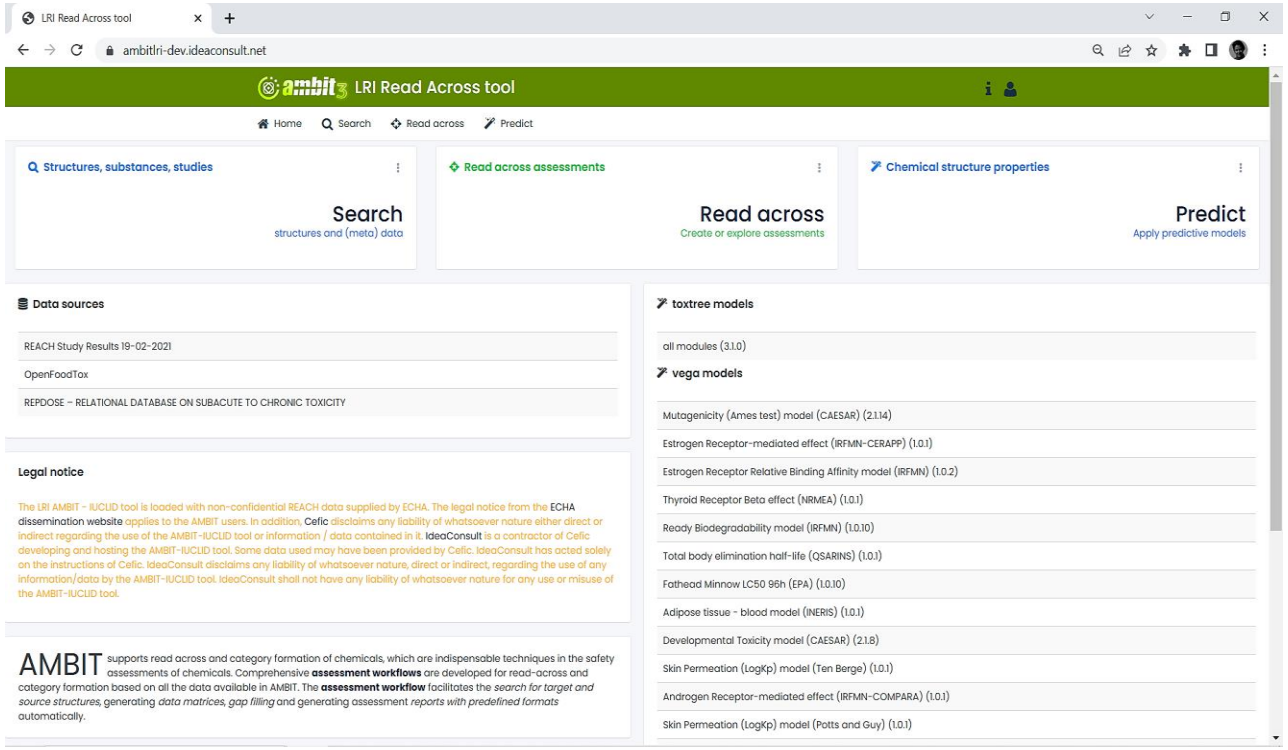
Substance similarity search, based on the maps (User Interface)

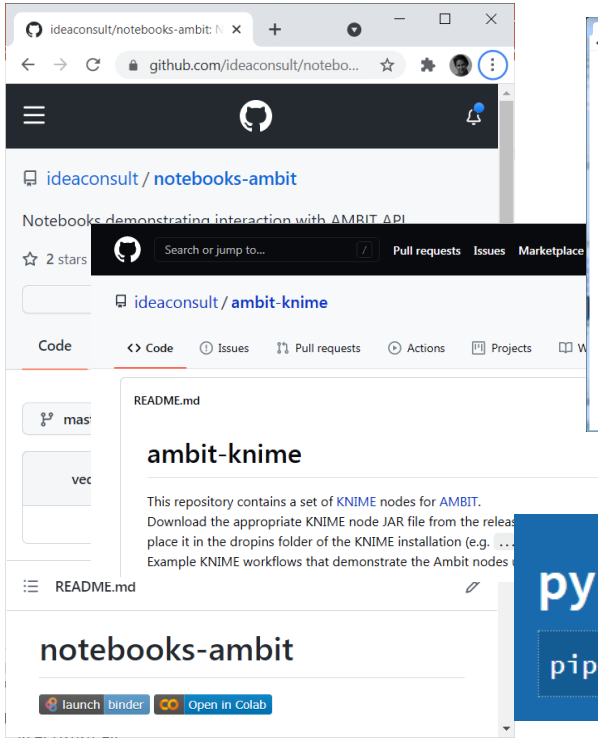


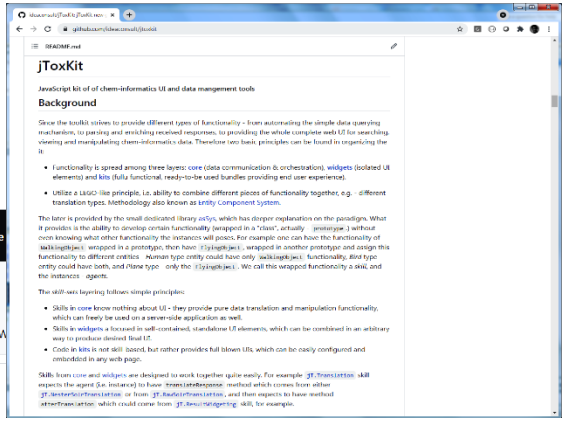
https://ambitlri-dev.ideaconsult.net



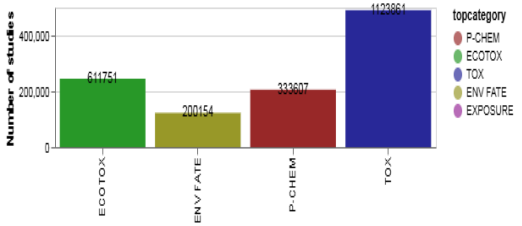
AMBIT 3 LRI Read Across Tool

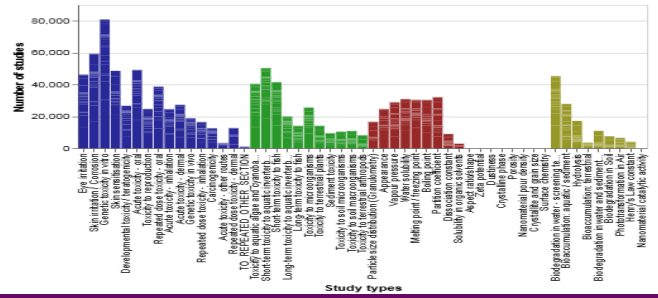






REST API & libraries



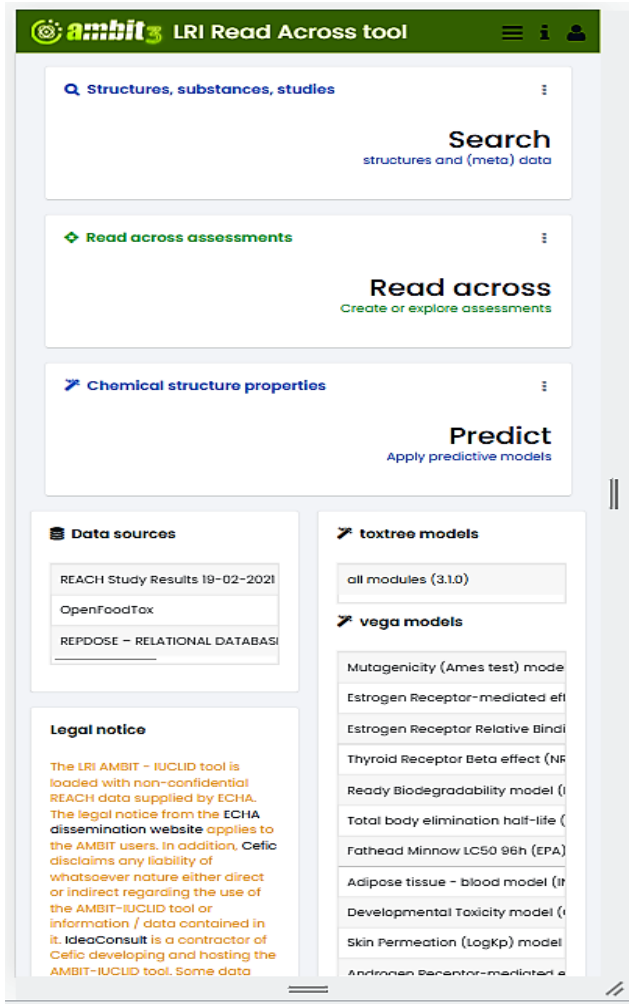


https://ambitlri-dev.ideaconsult.net

support@ideaconsult.net

AMBIT 3 LRI Read Across Tool

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



Frontend updates

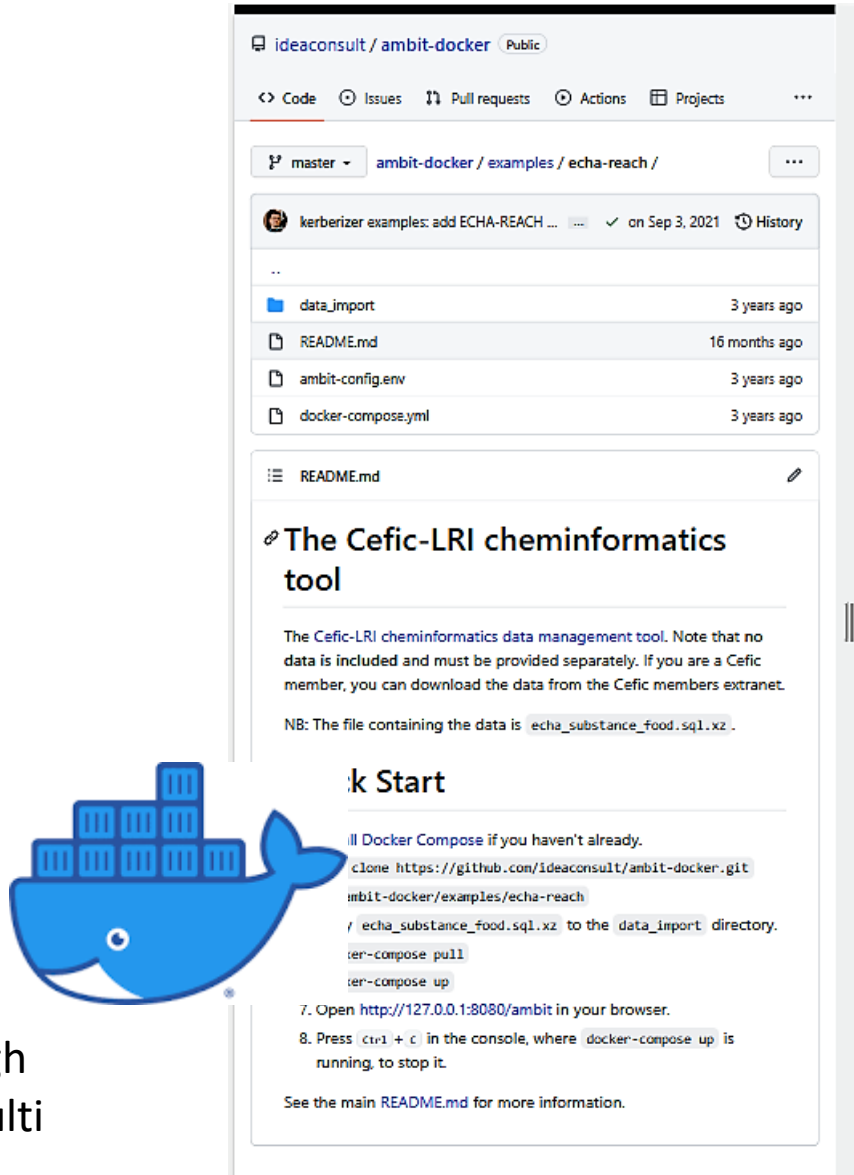
- New responsive design
- Upgrade of existing user interface
  - Read Across, Structure search
- New user interface
  - Dashboard, substance similarity

New functionalities

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

Backend updates:

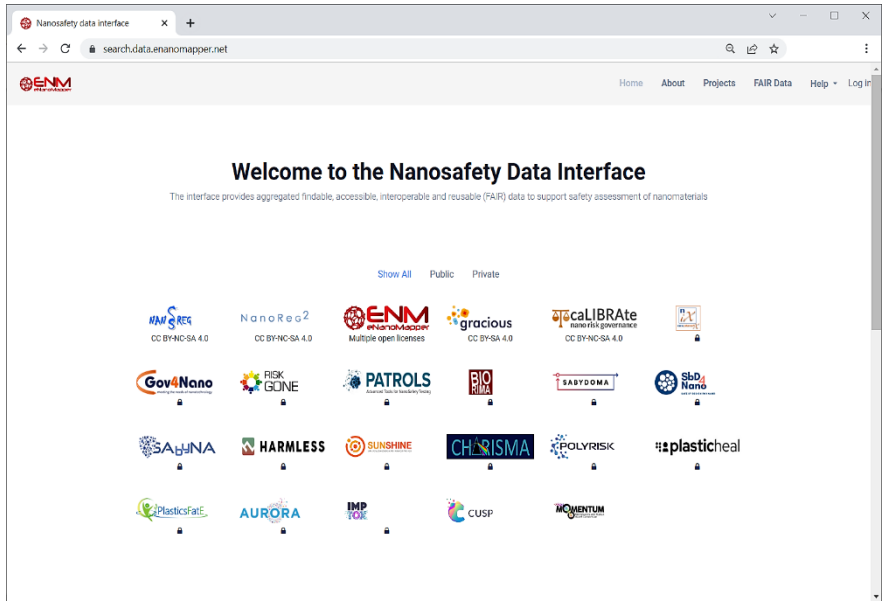
- New data integration approach through search engine (metadata, data and multi property similarity)
- Modular (microservices)

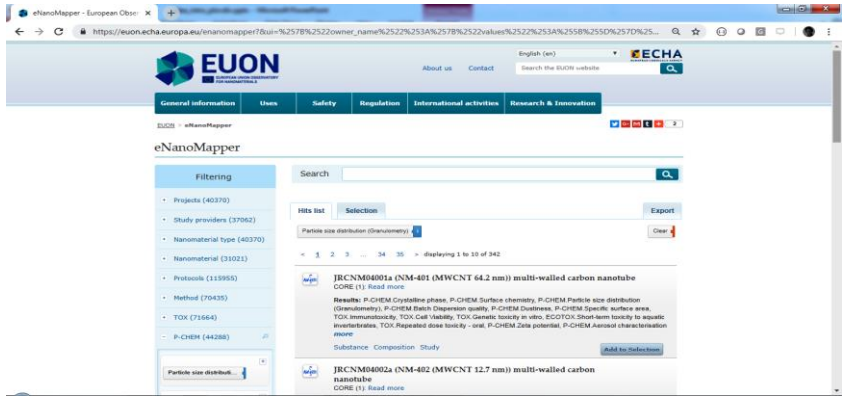


https://github.com/ideaconsult/ambit-docker/

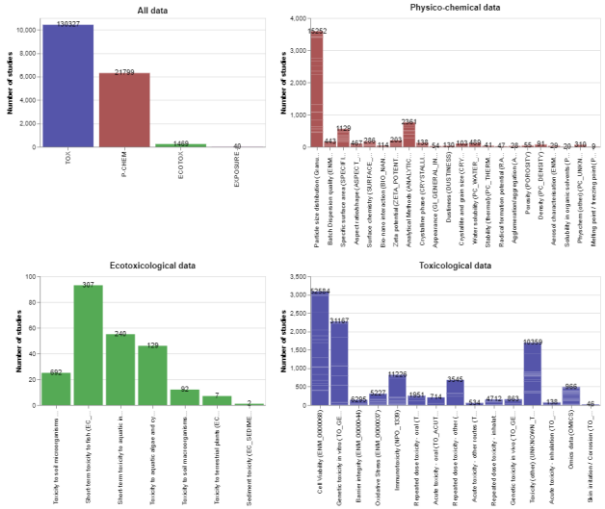
Distribution using modern containerized approach

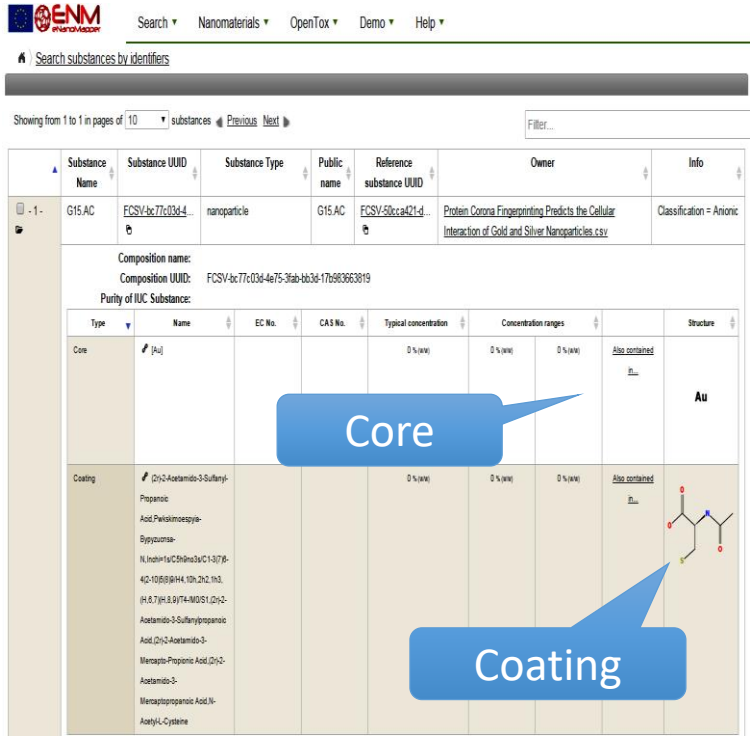
(nano/advanced)materials are substances





A large compilation of nanosafety data, generated by past and ongoing EU funded NanoSafety Cluster projects have been integrated.





Nanomaterials safety data (public & private)
https://search.data.enanomapper.net
EUON (public enanomapper)
https://euon.echa.europa.eu/enanomapper

nature nanotechnology

Analysis | Published: 20 May 2021

Towards FAIR nanosafety data

Nina Jeliazkova, Margarita D. Apostolova, [...]Penny Nymark

Nature Nanotechnology **16**, 644–654 (2021) | Cite this article

Thank you!

Questions?