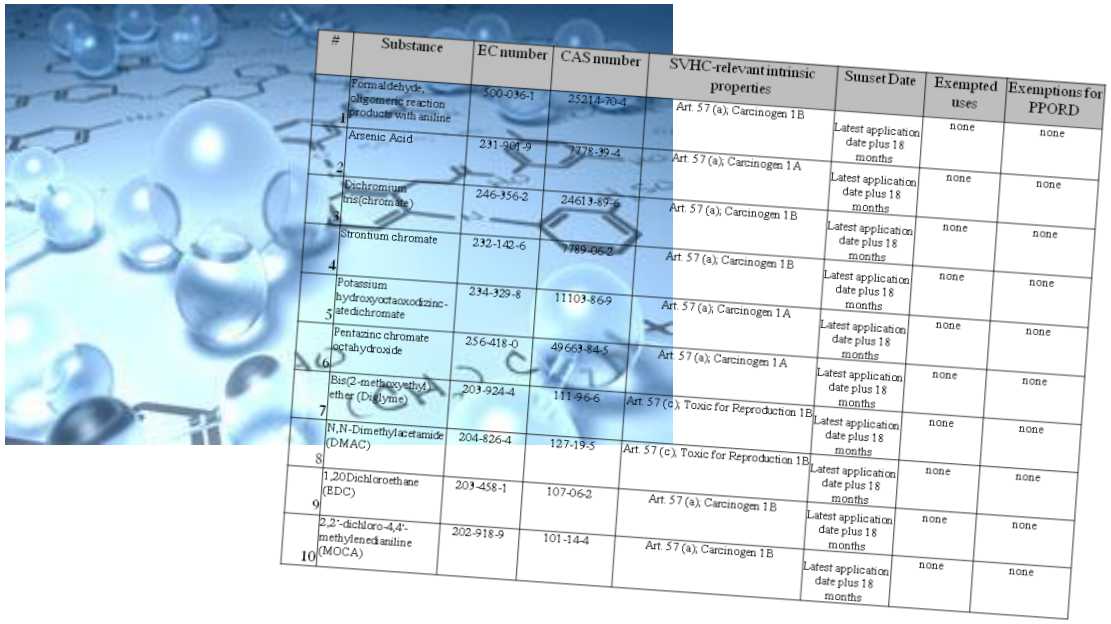
Selecting and curating REACH substance data

Nicolò Aurisano, Peter Fantke

Quantitative Sustainability Assessment, DTU Management
Technical University of Denmark



#	Substance	EC number	CAS number	SVHC-relevant intrinsic properties	Sunset Date	Exempted uses	Exemptions for PPORD
1	Formaldehyde, oligomeric reaction products with aniline	500-036-1	25214-70-4	Art. 57 (a); Carcinogen 1B	Latest application date plus 18 months	none	none
2	Arsenic Acid	231-901-9	7778-39-4	Art. 57 (a); Carcinogen 1A	Latest application date plus 18 months	none	none
3	Dichromium tris(chromate)	246-356-2	24613-89-6	Art. 57 (a); Carcinogen 1B	Latest application date plus 18 months	none	none
4	Strontium chromate	232-142-6	7789-06-2	Art. 57 (a); Carcinogen 1B	Latest application date plus 18 months	none	none
5	Potassium hydroxyoctaoxodizinc-atedichromate	234-329-8	11103-86-9	Art. 57 (a); Carcinogen 1A	Latest application date plus 18 months	none	none
6	Pentazinc chromate octahydroxide	256-418-0	49663-84-5	Art. 57 (a); Carcinogen 1A	Latest application date plus 18 months	none	none
7	Bis(2-methoxyethyl) ether (Diglyme)	203-924-4	111-96-6	Art. 57 (c); Toxic for Reproduction 1B	Latest application date plus 18 months	none	none
8	N,N-Dimethylacetamide (DMAC)	204-826-4	127-19-5	Art. 57 (c); Toxic for Reproduction 1B	Latest application date plus 18 months	none	none
9	1,20Dichloroethane (EDC)	203-458-1	107-06-2	Art. 57 (a); Carcinogen 1B	Latest application date plus 18 months	none	none
10	2,2'-dichloro-4,4'-methylenedianiline (MOCA)	202-918-9	101-14-4	Art. 57 (a); Carcinogen 1B	Latest application date plus 18 months	none	none

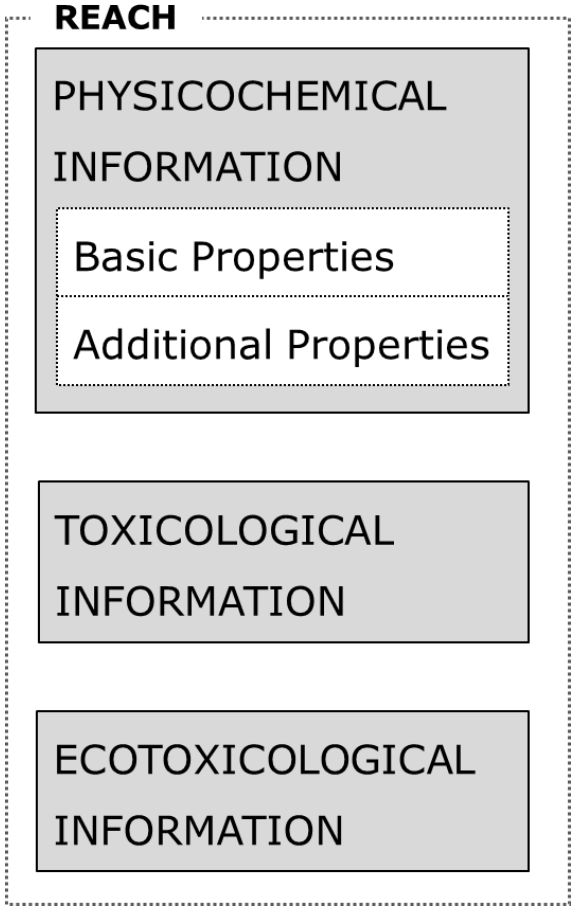
ISES Europe 2019 | 4-5 July 2019
Bilthoven, The Netherlands

Content

- REACH vs. data requirements in different contexts
- Data selection & curation method
- Application case study: Kow
- Conclusions & way forward

REACH vs. data requirements in different contexts

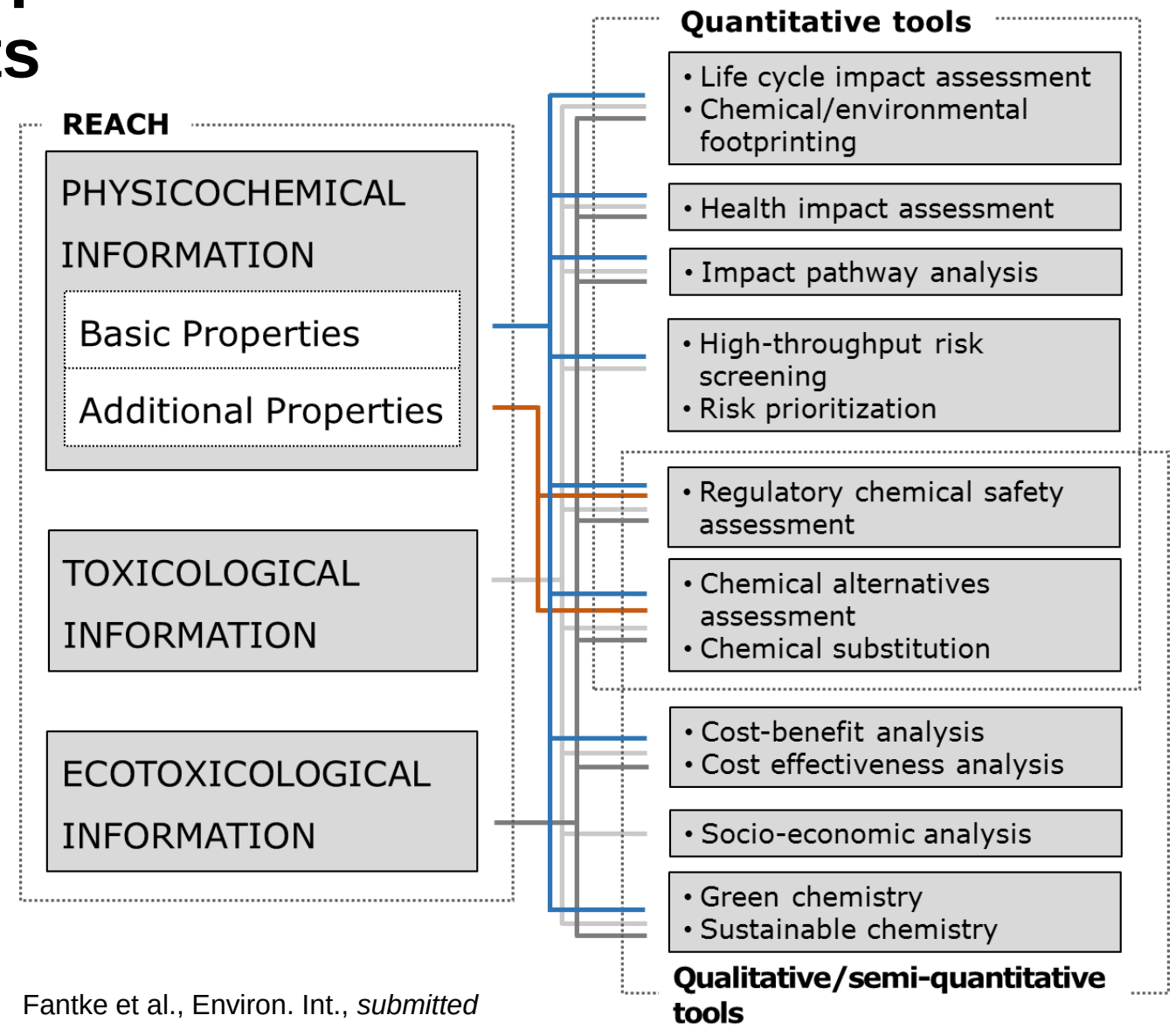
- Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH), Regulation (EC) 1907/2006
- Around 19.000 substances registered
- Level of information requirements based on yearly tonnage bands (ANNEX VII-X)



Fantke et al., Environ. Int., *submitted*

REACH vs. data requirements in different contexts

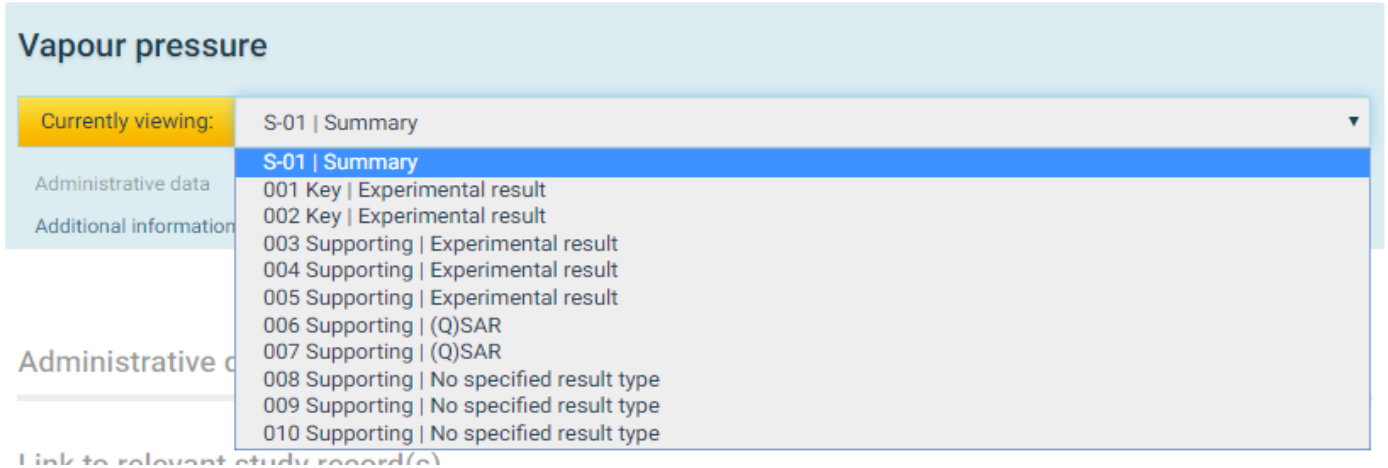
- Registration, Evaluation, Authorisation, and Restriction of Chemicals (REACH), Regulation (EC) 1907/2006
- Around 20.000 substances registered
- Level of information requirements based on yearly tonnage bands (ANNEX VII-X)



Fantke et al., Environ. Int., *submitted*

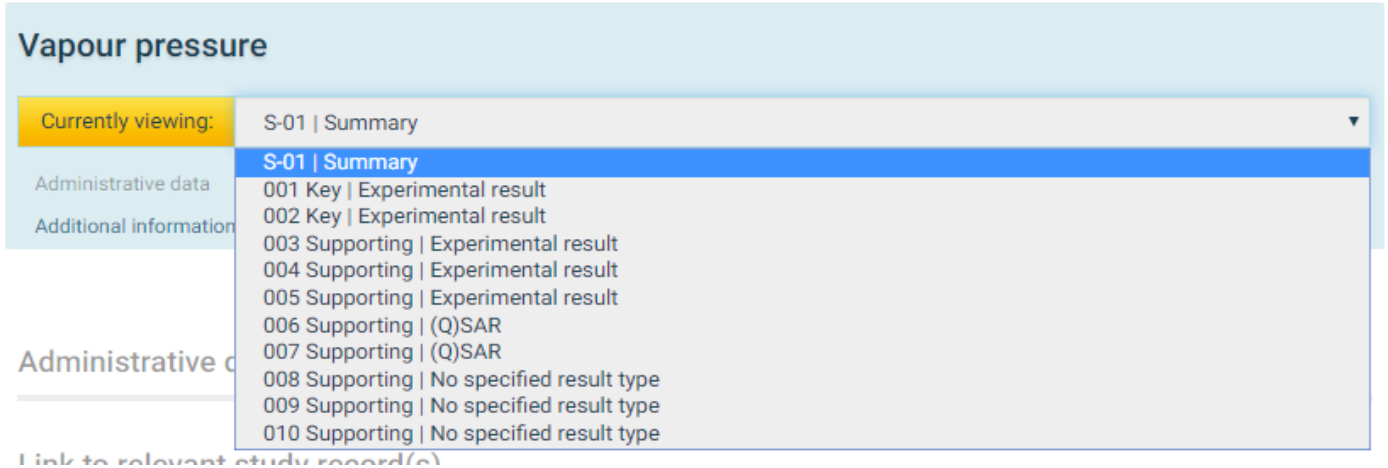
Using REACH data: Common issues

- Missing method for extracting, curating and selecting a single data point from the variety of reported data



Using REACH data: Common issues

- Missing method for extracting, curating and selecting a single data point from the variety of reported data
- Differences in data quality and underlying assumptions
- Considerable amount of critical information submitted via free text fields



Conclusions:	The 48-hour EC50 was 23.4 mg/L with a 95% confidence interval of 21.1 mg/L to 25.9 mg/L with a 95% confidence interval of 21.1 mg/L to 25.9 mg/L.
Conclusions:	After the first test without pH adjustment, the LC50 was 62.17 mg/L but severe pH effects were suspected after 1h of exposure. After the second test and pH adjustment, no mortalities were observed at the highest nominal concentration tested (215 mgL).
Conclusions:	After 96 hours of exposure to hexamethylene diamine, without pH adjustement, on Brachydanio rerio, a 96h-LC50 = 68 mg/L was populated.

Data selection and curation process (per substance)

Pre-processing:

- Test material and reported information check

Filtering:

- Filter application on available data

5 Steps data curation:

- Quality allocation
- Variability and Uncertainty assessment

Weighting:

- Derive one value across different data points

3 criteria implemented for assigning data quality and completeness:

- **Reliability**
- **Purpose Flag**
- **Type of Information**

Consider data quality in mean and uncertainty ranges

Expressed as squared geometric standard deviation (GSD^2):

Base Uncertainty: reflecting variability of the reported values

Criteria Uncertainty: reflecting assessed data points' quality

$$U_T = \mathrm{GSD}^2 = exp\left(\sqrt{(\ln U_b)^2 + \sum_i (\ln U_i)^2}\right)$$

Consider data quality in mean and uncertainty ranges

Expressed as squared geometric standard deviation (GSD²):

Base Uncertainty: reflecting variability of the reported values

Criteria Uncertainty: reflecting assessed data points' quality

$$U_T = \mathrm{GSD}^2 = exp\left(\sqrt{(\ln U_b)^2 + \sum_i (\ln U_i)^2}\right)$$

$$U_b = \sqrt{\frac{97.5^{th}\ \%ile}{2.5^{th}\ \%ile}}$$

$$\sum_i (\ln U_i)^2 = (\ln U_r)^2 + (\ln U_t)^2 + (\ln U_P)^2$$

$U_r = Uncertainty\ factors\ realibility$

$U_t = Uncertainty\ factor\ type\ of\ Information$

$U_P = Unceratnty\ factor\ Purpose\ flag$

Weighting: Combining data quality and variability

- Deriving a unique mean value and uncertainty around the mean per substance
- Taking into account both variability across data and quality per data point
- Data points with higher quality have higher influence on results

Weighting: Combining data quality and variability

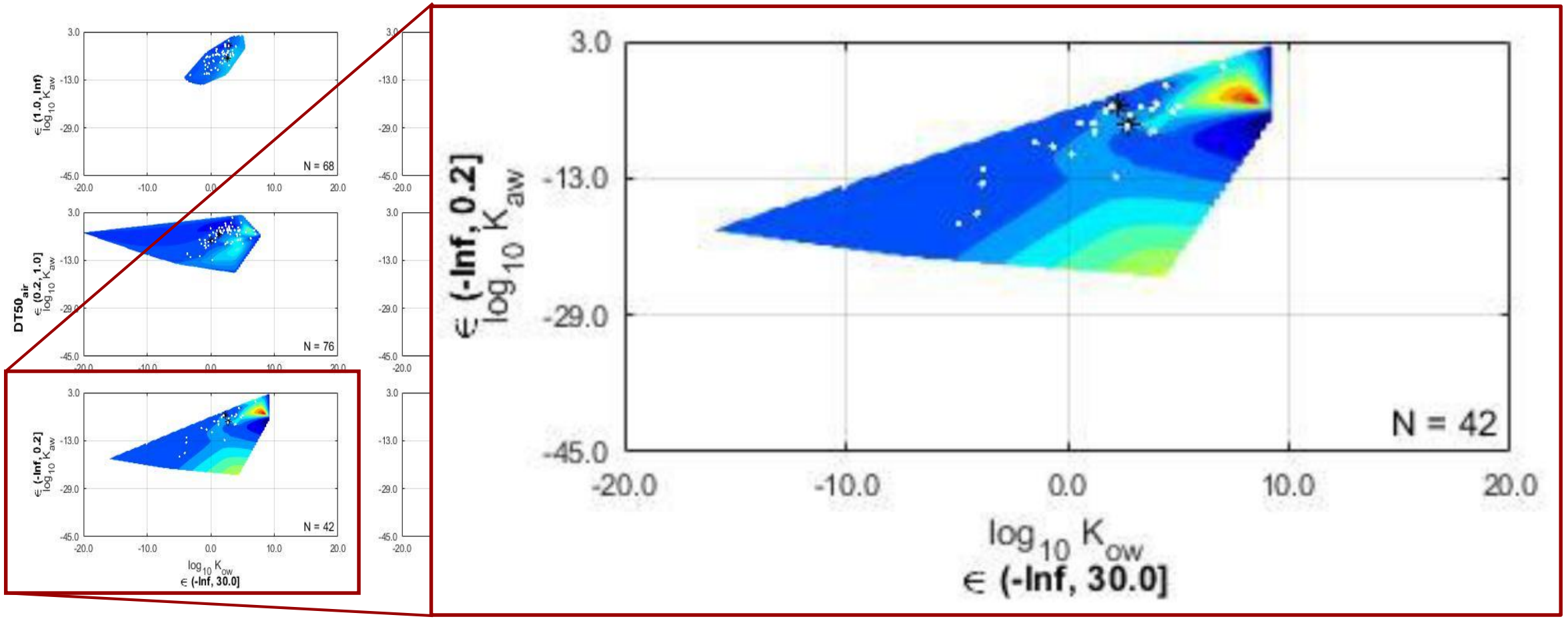
- Deriving a unique mean value and uncertainty around the mean per substance
- Taking into account both variability across data and quality per data point
- Data points with higher quality have higher influence on results

Data quality class	Normalized weights	Kow [l/l]	Weighted Kow	GSD^2	Weighted GSD^2
1 - high quality	0.901	1	**1.64**	11.7	**6.43**
2 - medium quality	0.096	10		18.9	
3 - low quality	0.002	100		58.4	

Application case study: Kow

- Focused on octanol-water partitioning coefficient (Kow)
- Selected **20 test substances** with criteria:
  - Cover broad spectrum of substance data space
  - Different REACH annual tonnage bands
  - Available in REACH and toxicity assessment model USEtox

Substance selection: Chemical data space



Conclusions and way forward

- REACH valuable data source, but **systematic data selection and curation needed** to be useful as input for different assessment frameworks (e.g. LCIA)
- We propose **semi-automated curation method** providing per substance **mean and uncertainty** ranges around the mean
- Next steps to **improving our method**:
  - Text mining (i.e. harvesting data from free text fields)
  - Add other quality criteria (e.g. test method and GLP)
  - Extend to cover other substance properties (e.g. vapour pressure, EC50)
  - Adapt to be applicable to other data sources (e.g. EFSA pesticide data)

Acknowledgements

- Jeroen Provoost and Panagiotis Karamertzani (REACH data)
- Michael Hauschild and Olivier Jolliet (input to curation method)

Nicolo Aurisano

PhD student

niau@dtu.dk
Produktionstorvet
Building 424
2800 Kgs. Lyngby
www.dtu.dk/english