Grouping of substances in screening

Round 4

ECHA Webinar – How are substances screened and shortlisted

14 February 2017

Dr Panagiotis (Panos) Karamertzanis
European Chemicals Agency

Content

- Substance grouping
  - why do we shift towards substance groups?
  - what do we mean with substance grouping?
  - methodology to construct substance groups
- How will groups be handled under common screening
- Conclusions

Common screening

Shortlisting substances – the approach so far

- ECHA in collaboration with MSCAs has developed a significant volume of screening scenarios to identify substances likely to be of concern

- The screening algorithms have been successfully used under common screening
  - common screening is in its 4th year of operation
  - ~200 substances manually screened each year
  - common screening succeeded earlier screening approaches that identified several hundred additional substances

- The screening scenarios cover both hazard and non-hazard (i.e. exposure) aspects and have been applied on individual substances that are prioritised regardless of their relationship to other substances in the registration database

- This approach has limitations

Common screening

Shortlisting individual substances – challenges

- For most substances there is some on going or completed regulatory activity either for the substance itself or for a *related* substance
- The outcome of manual screening is increasingly dependent on these other regulatory actions (conditional outcomes)
- For many of the substances not already short listed there are data deficiencies that pose difficulties in concluding manual screening
  - compliance check is an option, but not always necessary or the best solution

- From this round, some of the shortlisted substances are associated with other registered substances
- The idea is that once a substance is selected, then we also look for potentially related substances that may pose similar risks to human health or the environment
- Finding related substances at the stage of manual screening is not necessarily the same as fulfilling the criteria of Annex XI, 1.5 of REACH (grouping and read-across)

Common screening

Shortlisting related substances – benefits

- By pooling together all hazard information for related substances it may (to some extent) be possible to alleviate the difficulty to conclude manual screening due to missing information

- By looking at the whole group, including substances for which information generation is being considered or on going, it may be easier to fine tune our regulatory actions

- By considering related substances we can ensure consistency in our actions
  - act in a similar way when dealing with substance that pose similar risks

- There are efficiency gains if the learnings from assessing one substance can be applied to several others
  - capitalising the experience in assessing related substances has already been demonstrated in dossier evaluation

Substance grouping

General methodological aspects

Technically substance grouping for the purposes of common screening is based on

- method 1: structural similarity (using generated molecular structures), and
- method 2: proposed read-across and categories under REACH or other regulatory regimes

- There are other ways of grouping, such as
  - similar uses/common technical function
  - common mode of action
  - common metabolites, biodegradation or hydrolysis products

- ECHA and MSCAs are piloting such grouping methodologies

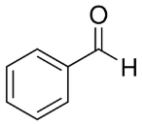
Substance grouping

Method 1: Structural similarity

- Molecular structures are "broken down" to functional groups taking into account connectivity up to a given distance ("chemical vocabulary")

- Every molecular structure is converted into a binary vector (vector with zeroes and ones)

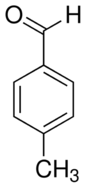
substance *i*
(registered)



$\text{bitvector } \boldsymbol{X}_i \in R^L$

11101010…00000001

substance *j*
(in the reference pool)



$\text{bitvector } \boldsymbol{X}_j \in R^L$

11101000…00000011

several aspects need to be considered, such as the length and density of the fingerprint, and the nature of the chemical vocabulary

- We compute the distance using a distance function (typically Tanimoto)

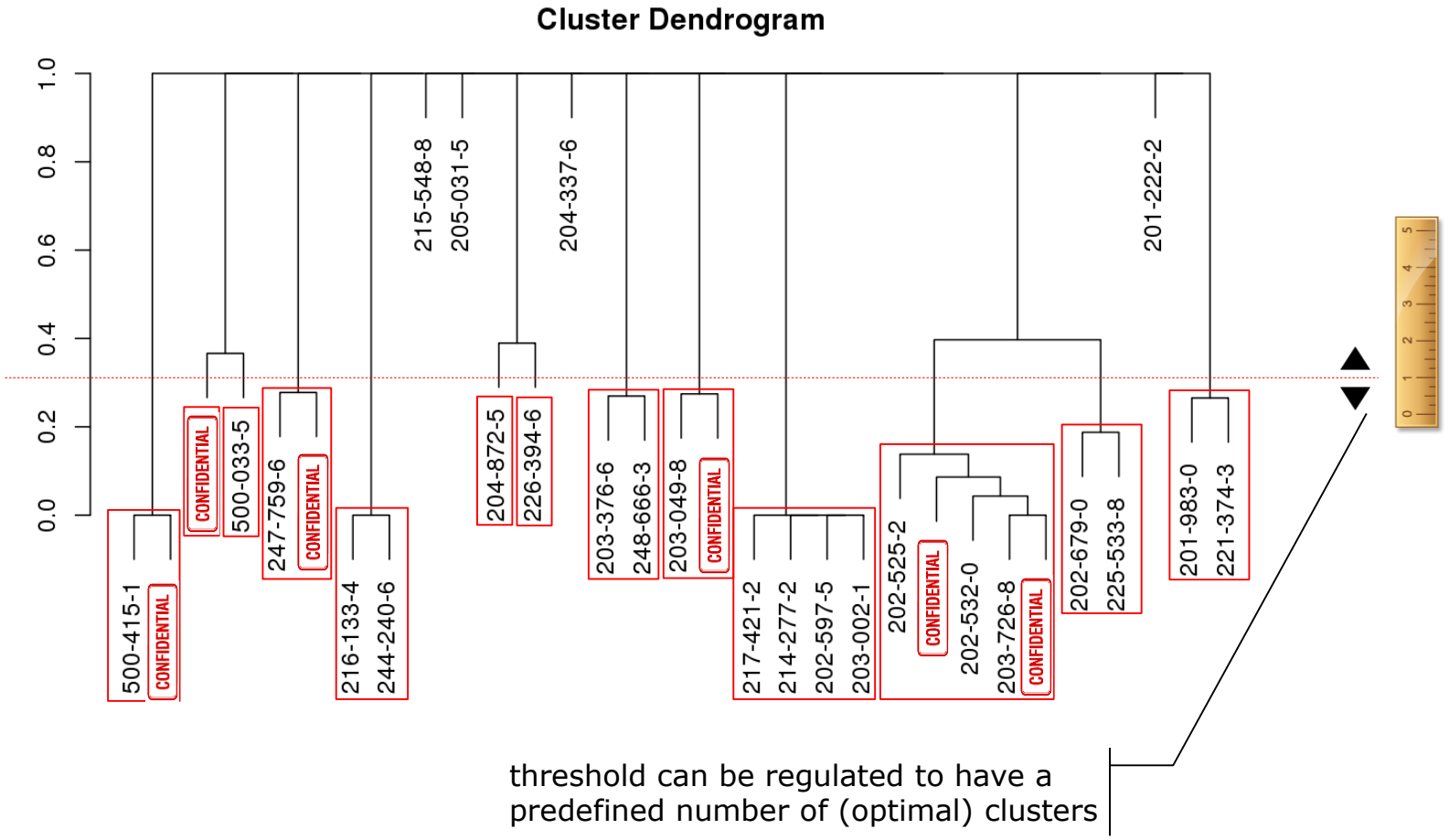
$$d_{ij} = T_{ij} = 1 - \frac{\boldsymbol{X}_i \cdot \boldsymbol{X}_j}{|\boldsymbol{X}_i|^2 + |\boldsymbol{X}_j|^2 - \boldsymbol{X}_i \cdot \boldsymbol{X}_j} \in [0,1]$$

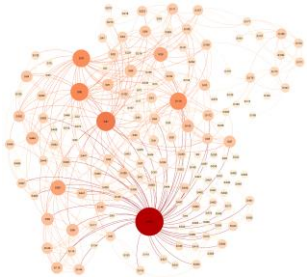
distance = 0 means identical structures

distance = 1 means completely different structures

Method 1: structural similarity

From structural distances to business useful information





Substance grouping

Method 2: Use of read-across/categories

- Related structures were identified by collecting analogues from one-to-one read-across or category statements proposed by either registrants or regulatory authorities
  - we can argue that the fact that a registrant or authority proposed a read-across or category has more significance for substance grouping than structural similarity alone
  - there should be argumentation that any differences in structure are not (eco)toxicologically important
  - the validity of read-across and category arguments can be examined during manual evaluation

- The following sources of analogues have been used (so far)
  - one-to-one read across arguments in the endpoint study records in the IUCLID dossiers of the parent substances
  - explicit categories in the IUCLID dossiers of the parent substances
  - (hidden categories in the IUCLID dossiers of the parent substances)
  - categories in the HPVIS programme of USEPA
  - NICNAS tier II human health categories, IMAP programme, Australia
  - OECD categories
  - …

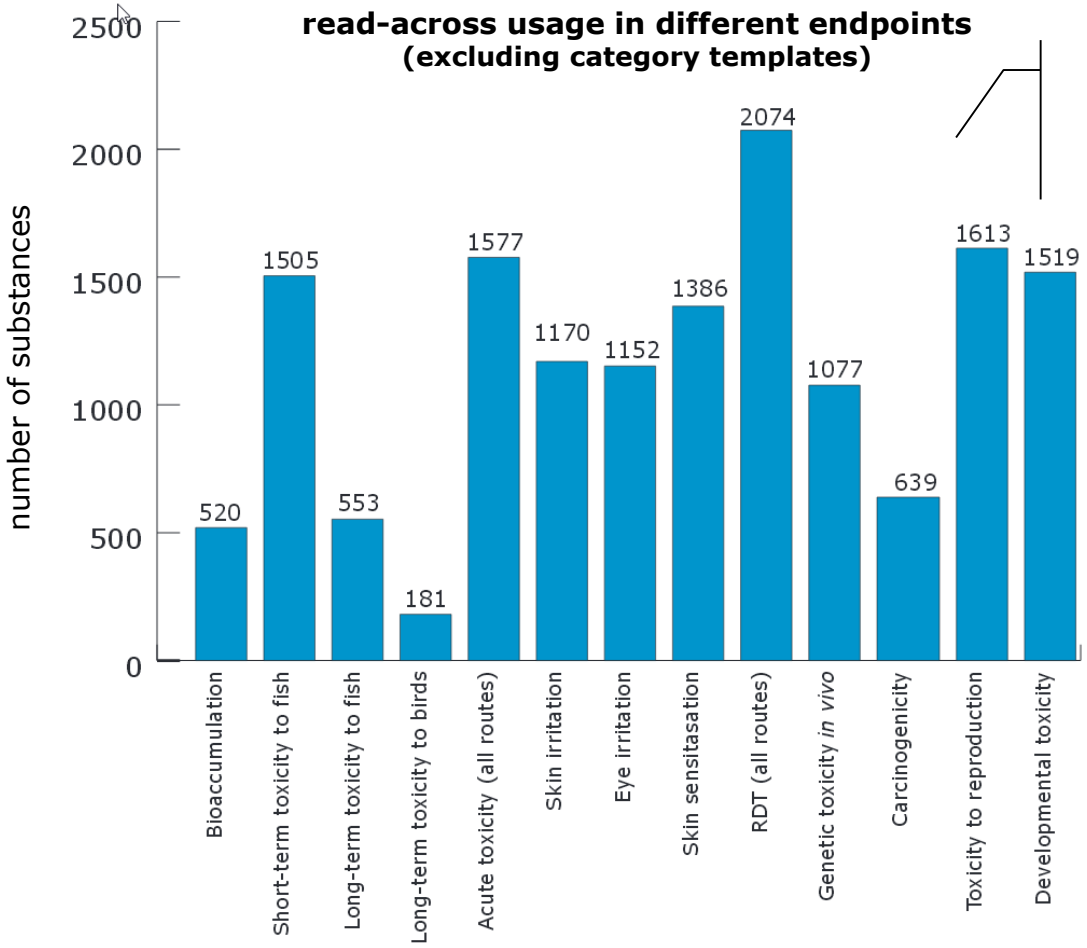
- The list of external sources will be extended further in the future

http://webnet.oecd.org/HPV/UI/ChemGroup.aspx
http://www.epa.gov/hpvis/
http://www.nicnas.gov.au/chemical-information/imap-assessments

Read-across and categories

Article 177(3) report

based on an analysis of 3813 lead and individual dossiers for **3663** substances at or above 100 ton/y



these numbers do not include read-across/category arguments in IUCLID category objects

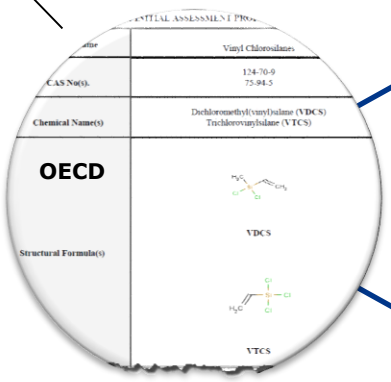
2014 Article 117(3) report
https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2014_en.pdf

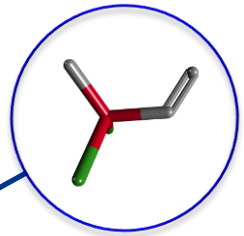
Method 2: use of read-across/categories

An example source of information: OECD categories

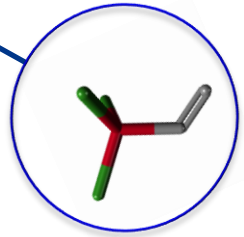
These two substances will be linked even if the registrations/notifications do not contain a read-across or category

vinyl chlorosilanes
(OECD category)





- CAS number 124-70-9
- Chemical name silane, dichloroethenylmethyl-
- *CAS number (alternative) 66062-55-3*



- CAS number 75-94-5
- Chemical name silane, trichloroethenyl-
- *CAS number (alternative) 127290-3-7-3*

We may argue that this relation would be captured with structural similarity too. Still linking the substances because of an OECD category gives different weight to the finding. Moreover, in other cases the structural relationship is not so obvious

OECD vinyl chlorosilane category
http://webnet.oecd.org/HPV/UI/ChemGroup.aspx#ctl00_ContentPlaceHolder1_156-header

Substance grouping

Putting everything together

- ECHA has developed an approach that allows combining the different grouping tools and methods into a single algorithm

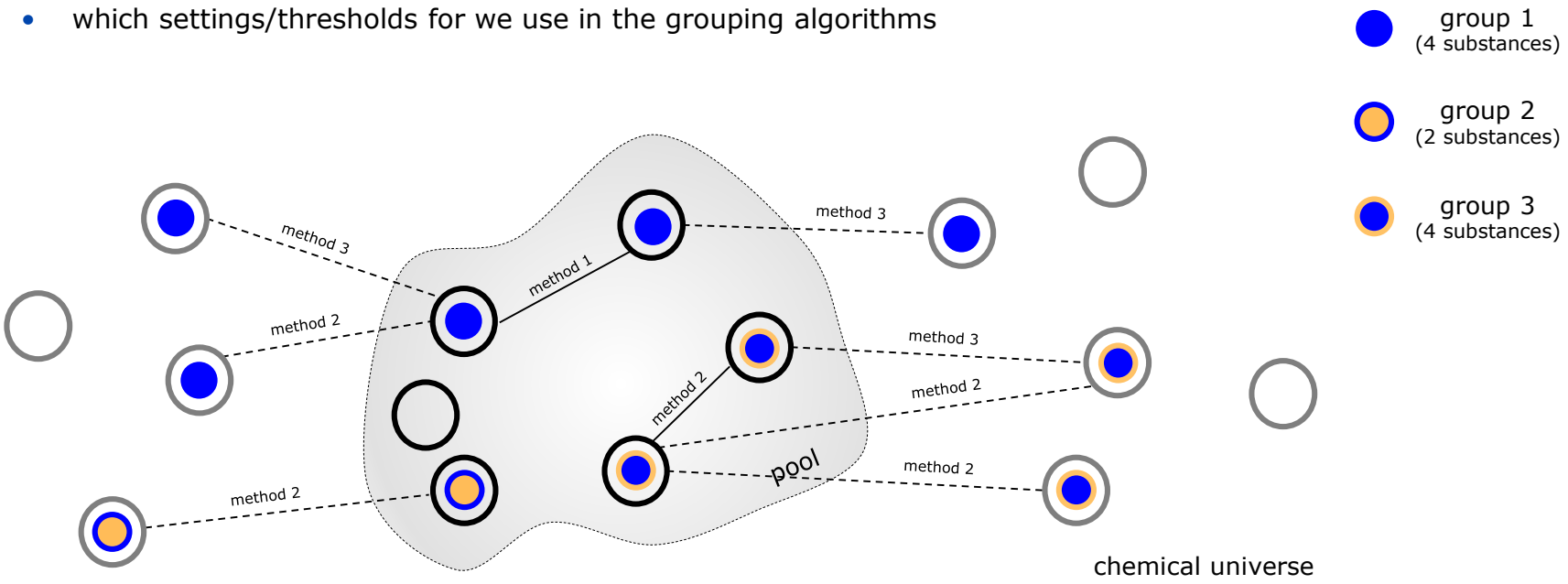
- This algorithm is heavily parameterised and the results we obtain depend strongly on the settings we use
  - we can vary the type and number of analogues
  - we can adjust how “tight” grouping should be

Substance grouping

Putting everything together

In order to understand the scope of grouping we need to define:

- which substances are in the pool we are trying to find analogues of
- which substances are in the chemical universe we are pulling analogues from
- which methodology do we use to do grouping
- which settings/thresholds for we use in the grouping algorithms



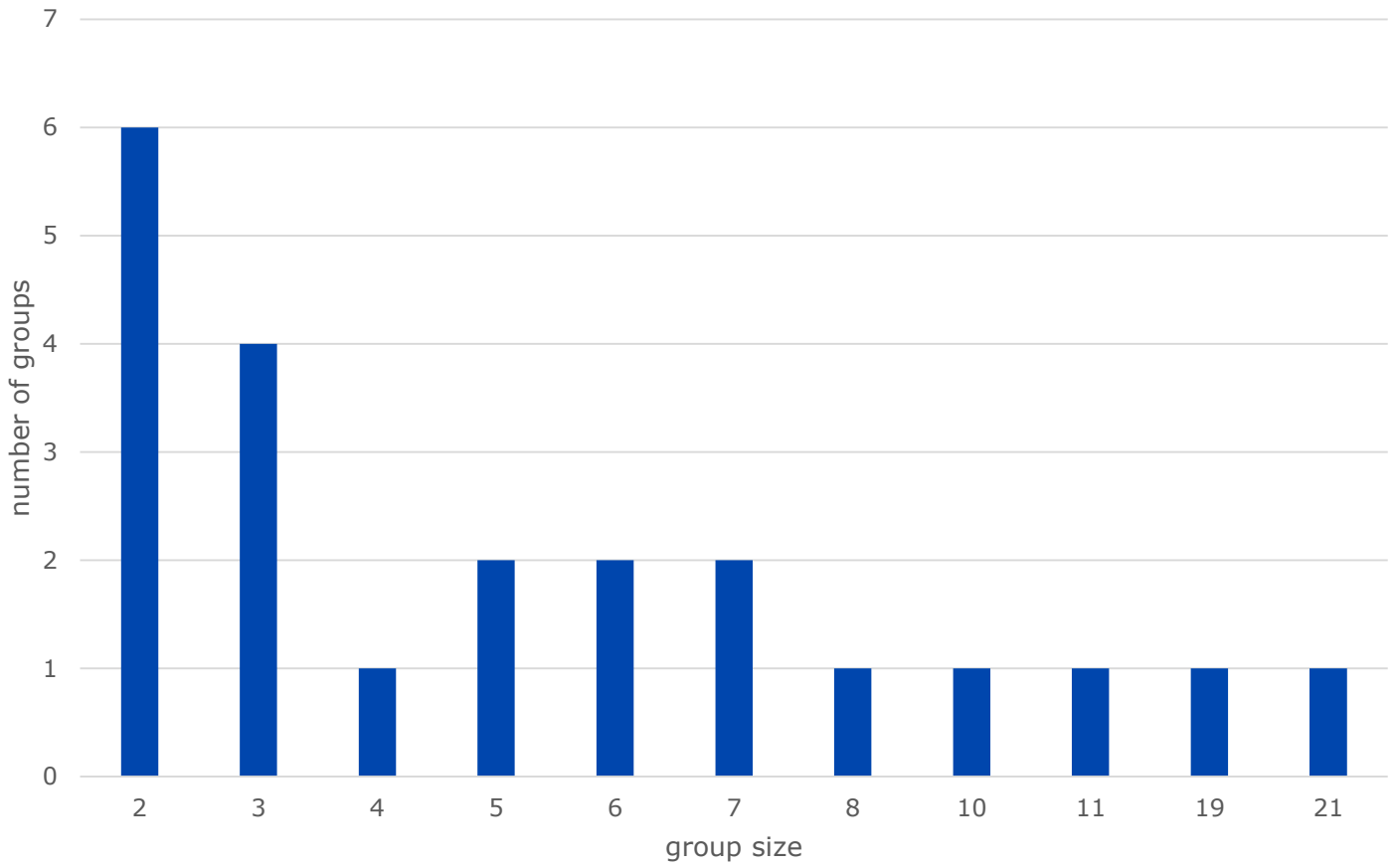
Substance grouping

How we found related substances this screening round

- We created groups for ~30 substances that were short listed using the hazard and non-hazard screening criteria; this led to the formation of 22 substance groups
- We also created two groups for substances in the candidate list as a pilot project to assess the possible substitution
  - ECHA and MSCAs are likely to intensify the formation of groups around substances that are already regulated in case there are similar hazards and there is possibility for unfortunate substitution
- We identified related substances using all mechanisms described earlier from the REACH registration database and C&L inventory (chemical universe)
  - the majority of substance linkages came from read-across statements in the registration dossiers
  - with regard to structural similarity we used strict distance thresholds
- The substance groups were manually examined by ECHA to correct a small number of inaccurate substance associations, e.g. due to erroneous identifiers
  - we did not evaluate the quality of the read-across or category justifications
  - a first assessment of the robustness of the substance group will take place during manual screening

Round 4 short list

Distribution of group sizes



ECHA
EUROPEAN CHEMICALS AGENCY

Substance grouping

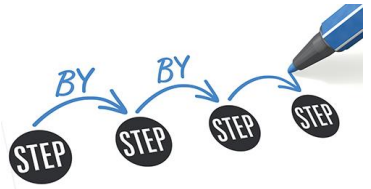
How are substance groups handled?

- Member State Competent Authorities (MSCAs) will manually screen the whole substance group and not individual substances within a group
- It is possible that more than one MSCA, and possibly ECHA, will collaborate when carrying out manual screening
- The validity of the group and its exact boundaries may be altered during manual screening
  - for example, the substances that have been related due to close structural similarity will be scrutinised to assess whether the hazard properties can be dissimilar despite the small structural distance
  - additional substances may be added to the group or the group may be split if the grouping approach is refined based on an enhanced understanding on the inherent properties of the substances as manual screening progresses
  - ECHA will remain in close collaboration with MSCAs and assist with the grouping as required
- For individual substances and small substance groups, Member State Competent Authorities will follow the general timelines of common screening, i.e.
  - if the out come is listing in CoRAP, end of May 2017
  - for all other outcomes, end of July 2017
- For larger groups the generic timelines for manual screening may be extended

Substance grouping

How are substance groups handled?

- MSCAs may
  - propose different manual screening outcomes for different substances in the group
  - decide that the same regulatory process is suitable for all substances in the group
  - or defer the assessment of some substances in the group to a later point in time, and after the generation of information for the remaining group members
- For larger, complex groups for which there has been prior regulatory activity, such as Dossier Evaluation or Risk Management Option Analysis, MSCAs and ECHA may also contact the registrants to discuss the identified concerns and possible further information needed to assess the whole group
- This is a pilot exercise and we will learn as we work. ECHA and MSCAs do not have all answers!



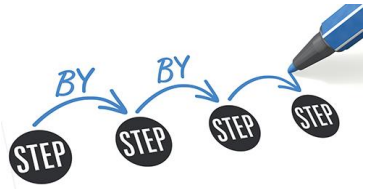
Substance grouping

What can registrants do?

- Read-across and categories are the most commonly used alternative approach to fulfil the information requirements

- The read-across and category arguments are used at face value by algorithms
  - if the quality of the read-across is poor we may pull together datasets of substances that do no behave similarly
  - when we associate substances we also pull together the hazard findings, that include external experimental data and predictions
  - hence, inclusion of unjustified read-across/category arguments do not necessarily make a stronger case
  - instead they may lead to the identification of additional and perhaps erroneous hazards that need to be followed with the registrant

Read-across assessment framework
https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across



Substance grouping

What can registrants do?

- What can the registrants do?
  - use read-across and category arguments wisely and adequately and appropriately document them
  - start with a hypothesis driven justification why data from one substance used for another for each endpoint
  - make use that the identity of all substances used is clear to avoid unintended substance associations
  - explain how structural similarity and dissimilarity affect the predictions
  - toxicokinetic information can considerably strengthen the robustness of the read-across
  - reliable and adequate mechanistic and "omics" data can be beneficial
  - unsubstantiated arguments of the type "substances are similarly metabolised" are not sufficient to justify the read-across but they trigger our algorithms to pull together hazard datasets

- ECHA's Read-across Assessment Framework structures the scientific evaluation of grouping and read-across in REACH (RAAF)
  - RAAF summarises the assessment considerations for each read-across approach for human health
  - in case the assessment conclusion of your read-across is negative, you may want to re-examine the usefulness of the read-across

- Watch out for the planned RAAF extensions (environmental endpoints, multi-constituents and UVCBs)

Read-across assessment framework
https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across

Read-across and categories

QSAR Toolbox

- Co-developed by ECHA and OECD
- Free to download[1]
- It provides
  - available experimental data
  - endpoint specific information of possible mode of actions based on the structure
  - support in building categories and read-across cases based on structural and mechanistic similarity
- Guidance and practical examples on how to use the QSAR Toolbox is available on ECHA website[2]

[1] https://www.qsartoolbox.org/
[2] https://echa.europa.eu/support/oecd-qsar-toolbox

These are general principles. There are differences between the category and analogue approaches. Please consult the IUCLID 6 manual!

Read-across and categories in registration dossiers

Transition to IUCLID 6

- In case you decide to update your dossier, please note that the information with regard to read-across has been restructured in IUCLID

- In IUCLID 6, reporting of analogue read-across to be done with **two** records: **source** and **target**

- Source record is the experimental study on source material
  (Type of information = experimental study)

- Target record is the outcome of applying read-across
  (Type of information = read-across)

- Target records are only checked for information relevant to read-across outcome:
  - administrative data, justification
  - target material
  - result

- Migration will keep existing read-across endpoint study records as they are and completeness check will not enforce their reworking. But recommendable to revise when feasible.

	Relevant information	
Endpoint study record	Source record	Target record
Administrative data	X	X
Endpoint	X	X
Type of information	X	X
Adequacy of study	X	X
Robust study summary Used for classification Used for SDS	X	X
Study period	X	-
Reliability	X	-
Rationale for reliability incl. deficiencies	X	-
Data waiving	-	-
Justification for data waiving	-	-
Justification for type of information	X	X
Attached justification	X	X
Cross-reference	X	X
Data source	X	-
Materials and methods	X	-
Test materials	X	X
Result and discussion	X	X
Overall remarks, attachments	X	X
Applicant's summary and conclusion	X	X

*X = relevant chapters/fields; **X** = subject to completeness check*

Thank you

Panos Karamertzanis
screening@echa.europa.eu

Subscribe to our news at
echa.europa.eu/subscribe

Follow us on Twitter
@EU_ECHA

Follow us on Facebook
Facebook.com/EUECHA