Dr Panagiotis Karamertzanis

European Chemicals Agency, Finland

A common screening approach for REACH and CLP processes

Topical Scientific Workshop on New Approach Methodologies in Regulatory Science

20 April 2016

Dr Panagiotis (Panos) Karamertzanis
European Chemicals Agency

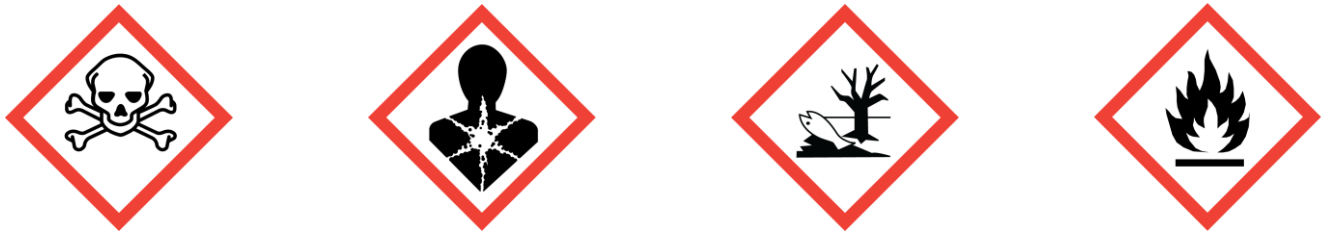
Content

- Background
  - definitions and timelines
- Tools and methods
  - underlying data used for screening
  - generation of molecular structures
  - overview of hazard scenarios
  - use and exposure
  - substance grouping
- Results obtained so far
  - outcome of rounds 2 and 3
  - progress with on-going round 3
- Conclusions

Common screening approach

The aim of common screening is to identify and prioritise those substances for which regulatory action is likely to have a significant positive impact on the protection of human health or the environment

Substances that matter most



Common screening

soft measures, e.g. letter campaigns, Article 36 letters

Use of **all available data**

Allocate identified substances to the appropriate process (**if any**):

Generation of further information

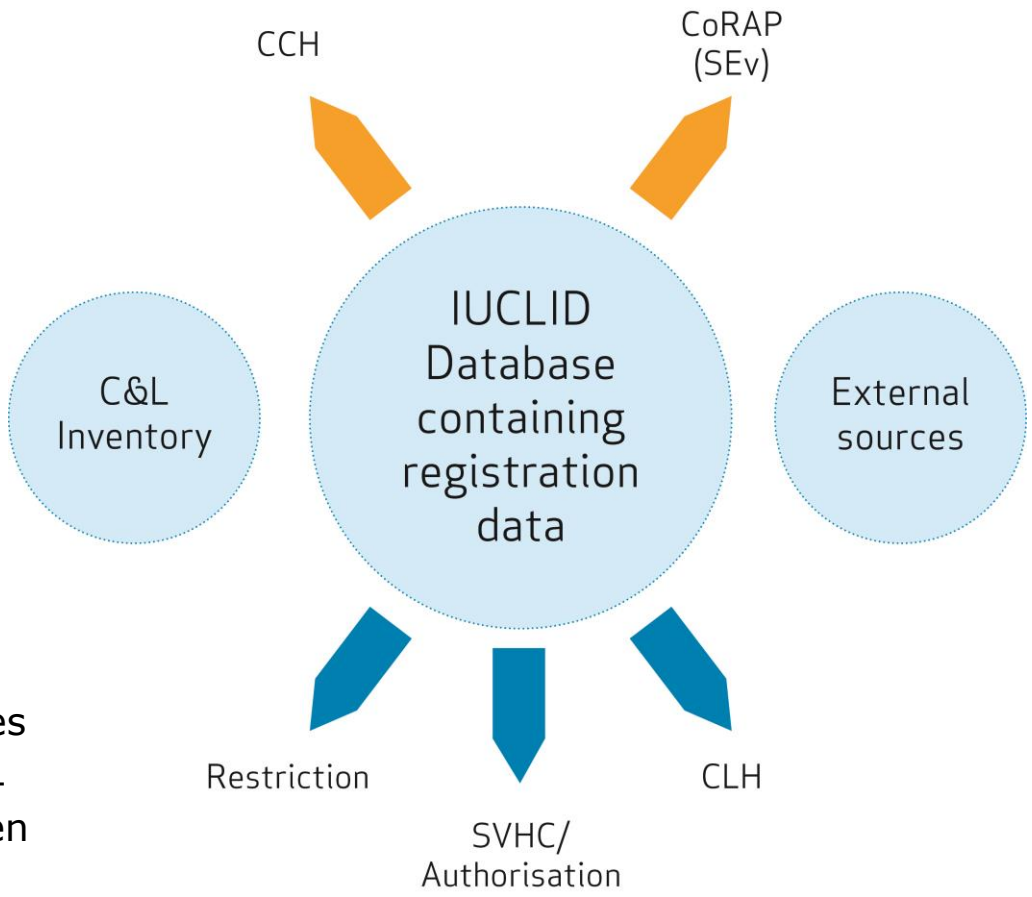
- Substance evaluation (SEv)
- Compliance check (CCH)

Regulatory risk management

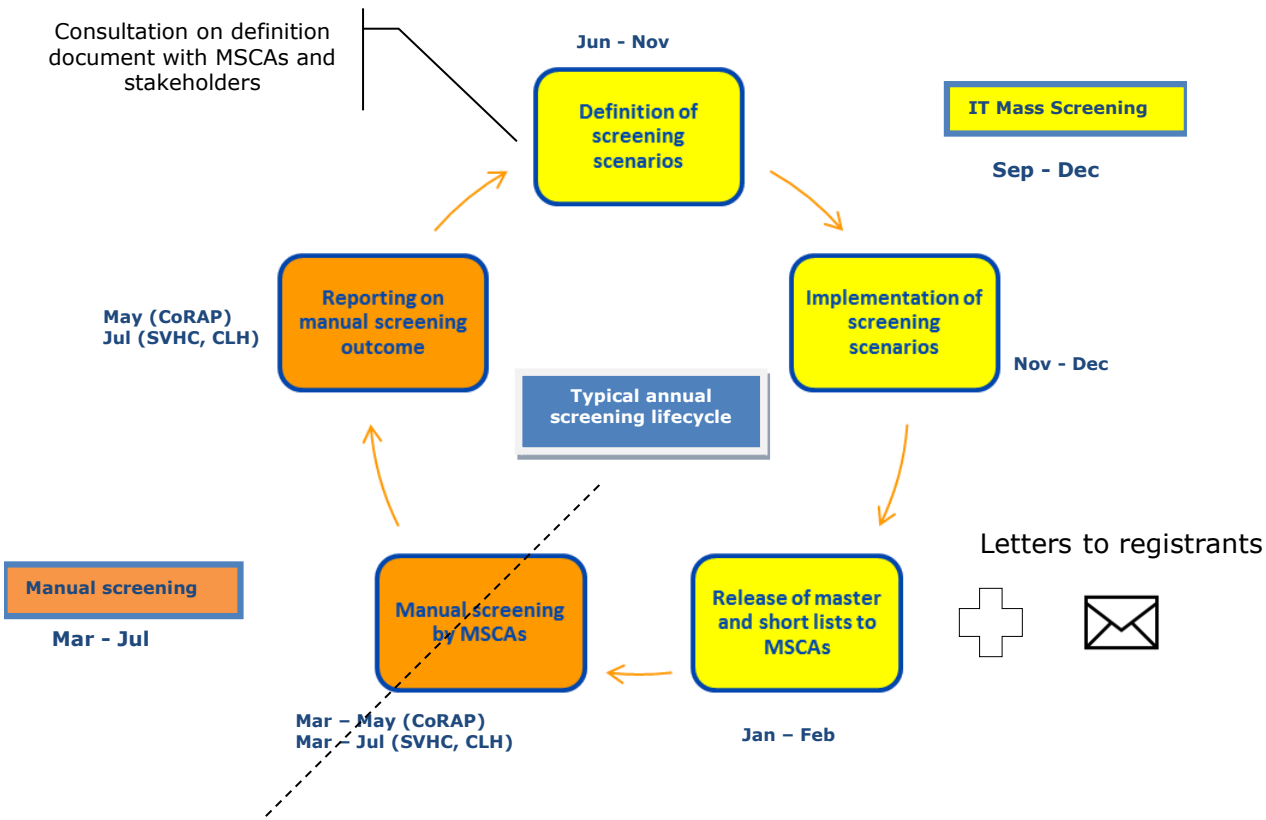
- Harmonised classification and labelling (CLH)
- Identification of SVHCs (possibly leading to authorisation)
- Restriction

Fully integrated approach:

- Optimal use of resources
- Avoids parallel processing of substances
- Ensures that the most effective regulatory option for each substance is chosen



Common screening
Yearly cycle

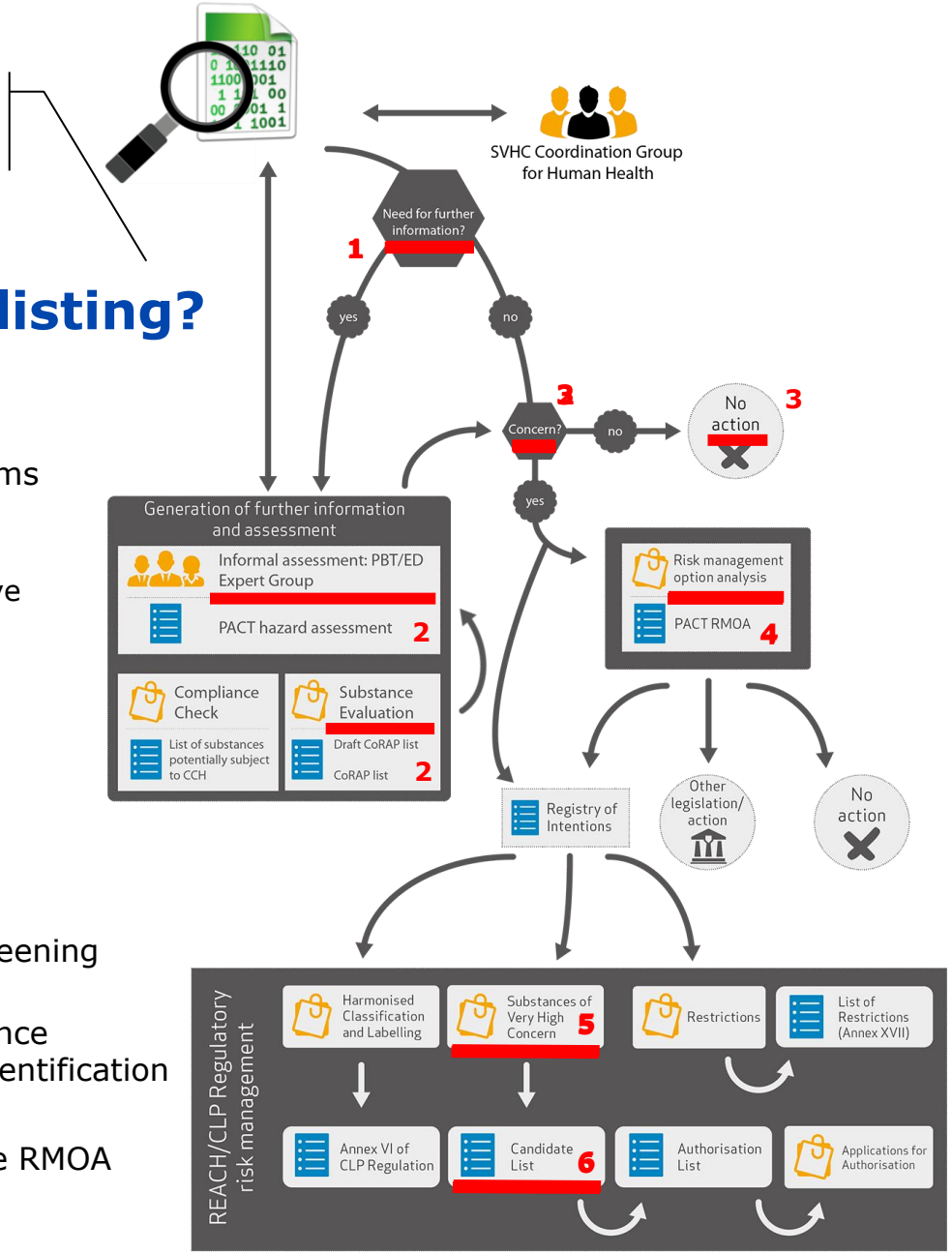


- IT screening results in a shortlist of substances
  - 200-300 substances annually
  - new information received, scenarios refined
  - 2 000-3 000 automated letters sent to registrants (reasoning for shortlisting and next steps)

- Member States select substances for further scrutiny from the short list
  - manual verification of IT screening outcome
  - holistic evaluation of substance
  - determine whether further regulatory action is required
  - feedback into IT screening to improve the process

Common screening
What happens after shortlisting?

- Common screening[1] uses automated algorithms to add substances in the shortlist for manual screening
  - including an indicative concern and an indicative regulatory process
- Member State experts manually evaluate the algorithm findings
  - shallow but wide evaluation
  - indicative concern/process are confirmed or modified
- Several options are possible once manual screening is concluded, e.g.
  - immediately next regulatory step is substance evaluation, followed by RMOA and SVHC identification
  - no action
  - input from the PBT/ED Expert Group before RMOA



[1] Common screening web pages
http://echa.europa.eu/addressing-chemicals-of-concern/substances-of-potential-concern/screening

Content

- Background
  - definitions and timelines
- Tools and methods
  - underlying data used for screening
  - generation of molecular structures
  - overview of hazard scenarios
  - use and exposure
  - substance grouping
- Results obtained so far
  - outcome of rounds 2 and 3
  - progress with on-going round 3
- Conclusions

Data used for screening

Information sources and data organisation

the master is the source of information for most screening activities

selection of substance (short listing criteria) → short list

algorithm output → master list

intermediate results

- initial pool of substances
- administrative warnings
- hazard scenario results
- non-hazard scenario results
- exposure indicators

sources of information

- registration data
  - ~50 000 registrations
- C&L notification data
  - ~500 000 notifications
- external data
  - experimental datasets (*in vivo* and *in vitro*)
  - use and exposure information
  - international regulatory output
  - NGO priority lists
  - structure information
  - …
- predictive methods
  - structural alerts
  - QSAR models
  - known metabolism paths and predictions

ECHA
EUROPEAN CHEMICALS AGENCY

Data used in common screening

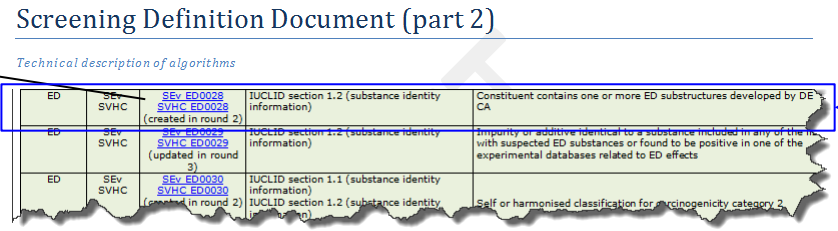
Structuring of results in the master list

once the query is executed a report is created that can be exported or further filtered

most screening objects have a code and an explanatory message

EC number	substance name	code	message
	CONFIDENTIAL CONFIDENTIAL	SEv ED 0025	the impurity (SMILES: [redacted]) with reference substance UUID "SNIF-4ae73e8f-eb83-30e2-abfa-ae1a05bcf8b7/ IUC5-647950e7-8146-4d63-b101-81de516dd5cd" and reference substance name "[redacted]" can be found at significant concentration (typical concentration: ca.1.0 % (w/w), concentration range: ca.0.5 ca.2.0 % (w/w)) in the composition with composition name "[redacted]" and composition local UUID "L-ca7eb4b7-a945-3f48-8a0f-7bc5789d8ad2" and is structurally identical to the suspected endocrine disruptor "suspected endocrine disruptor (TedX list) 120" in the TEDX list {structural search (matched identifier of type ref_sub_name),structural search (matched identifier of type iupacname),structural search (matched identifier of type i5_cas),structural search (matched identifier of type alternativecasname (provided by acs, starting from i5_cas))}
	CONFIDENTIAL CONFIDENTIAL	SEv ED 0025	the impurity (SMILES: structure could not be converted to smiles) with reference substance UUID "SNIF-4ae73e8f-eb83-30e2-abfa-ae1a05bcf8b7/ IUC5-647950e7-8146-4d63-b101-81de516dd5cd" and reference substance name "[redacted]" can be found at significant concentration (typical concentration: ca.1.0 % (w/w), concentration range: ca.0.5 ca.2.0 % (w/w)) in the composition with composition name "[redacted]" and composition local UUID "L-ca7eb4b7-a945-3f48-8a0f-7bc5789d8ad2" and is structurally identical to the suspected endocrine disruptor "suspected endocrine disruptor (TedX list) 120" in the TEDX list {numerical search (matched identifier of type i5_cas)}
	CONFIDENTIAL CONFIDENTIAL	SEv ED 0027	the reference substance (SMILES: [redacted]) matches the structural alert for endocrine disruption "DE_ED_structuralAlert 18" {structural search (matched identifier of type synonymname),structural search (matched identifier of type i5_casname),structural search (matched identifier of type i5_cas),structural search (matched identifier of type alternativecasname (provided by acs, starting from i5_cas)),structural search (matched identifier of type smiles),structural search (matched identifier of type iupacname),structural search (matched identifier of type ref_sub_name)}
	CONFIDENTIAL CONFIDENTIAL	SEv ED 0028	the constituent (SMILES: [redacted]) with reference substance UUID "SNIF-32129e9c-64bd-3ad5-b764-10612e60e7c2/ IUC5-647950e7-8146-4d63-b101-81de516dd5cd" and reference substance name "[redacted]" can be found at significant concentration (typical concentration: 97.0 % (w/w), concentration range: 95.0 99.0 % (w/w)) in the composition with composition name "[redacted]" and composition local UUID "L-ca7eb4b7-a945-3f48-8a0f-7bc5789d8ad2" and matches the structural alert for endocrine disruption "DE_ED_structuralAlert 18" {structural search (matched identifier of type synonymname),structural search (matched identifier of type i5_casname),structural search (matched identifier of type i5_cas),structural search (matched identifier of type alternativecasname (provided by acs, starting from i5_cas)),structural search (matched identifier of type smiles),structural search (matched identifier of type ref_sub_name),structural search (matched identifier of type iupacname)}
			the reference substance activated the "rtER Expert System ver.1 - USEPA" profiler for estrogen receptor binding as implemented in the QSAR

part 2 of the definition document contains browsable technical descriptions of all objects


Screening Definition Document (part 2)

Technical description of algorithms

ED	SEv SVHC	SEv ED0028 SVHC ED0028 (created in round 2)	IUCLID section 1.2 (substance identity information)	Constituent contains one or more ED substructures developed by DE CA
ED	SEv SVHC	SEv ED0029 SVHC ED0029 (updated in round 3)	IUCLID section 1.2 (substance identity information)	Impurity or additive identical to a substance included in any of the li with suspected ED substances or found to be positive in one of the experimental databases related to ED effects
ED	SEv SVHC	SEv ED0030 SVHC ED0030 (created in round 2)	IUCLID section 1.1 (substance identity information) IUCLID section 1.2 (substance identity	Self or harmonised classification for carcinogenicity category 2,

Generation of molecular structures

Important but challenging…

- The substance identity information in registrations is of paramount importance
  - allows links to external sources
  - allows use of predictive models
  - allows substance grouping

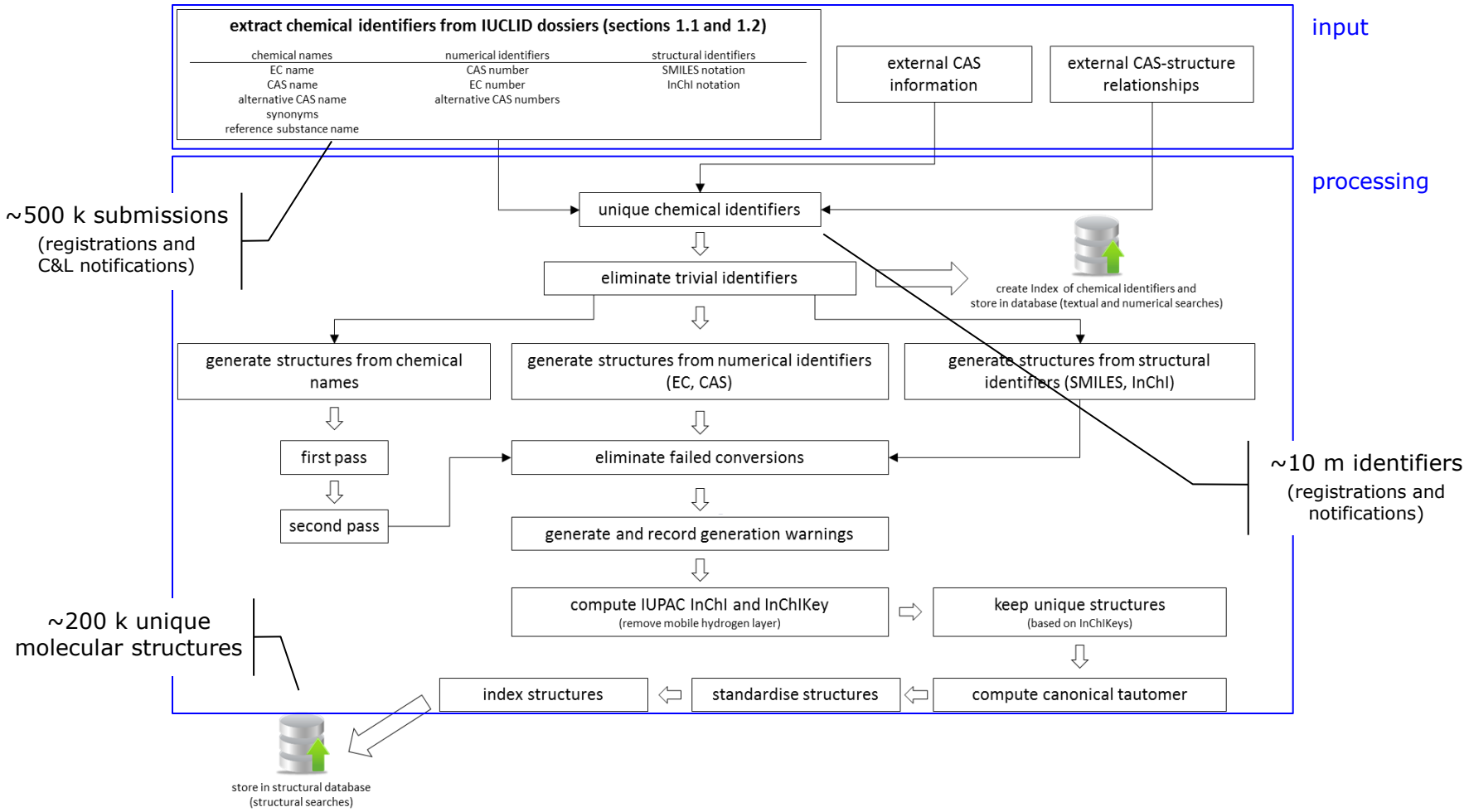
- Still, substances very often cannot be represented with a single, discrete molecular structure
  - multiple registrants for the same substance
  - multiple compositions
  - presence of several constituents, impurities or additives in the same composition
  - UVCBs (generic reference substances)
  - multiple identifiers provided for the same constituent, impurity or additive
  - multiple possible representations of the same structure (tautomers, salts, nitro groups…)

- Common screening streamlined the use of all available substance identity information in registrations, C&L notifications and external sources

Basic requirement is to exploit to the maximum the structuring of information offered by IUCLID

Generation of molecular structures

Machinery



Overview of hazard scenarios
Endocrine disruptors

Screening is exhaustive; however individual findings (e.g. an impurity at low concentration in a low volume substance) are often not sufficient to require regulatory action

- Endocrine disruptors
  - constituents, impurities or additives in regulatory lists for ED properties (identical or similar)
  - constituents, impurities or additives in ED priority lists by NGO or others, such as TEDX, WHO, SIN (identical or similar)
  - evidence of toxicity to endocrine organs in registration dossiers (using both structured and unstructured information)
  - PPP and biocide surrogate criteria (Carc. 2 & Repr. 2 self and harmonised classifications)
  - structural alerts and QSAR models
  - ToxCast ED related assays
  - constituents, impurities or additives in training sets of ED QSAR models (identical or similar)
  - scenarios categorised according to the mode of action when possible (estrogenic, androgenic, thyroid)

Overview of hazard scenarios
Environment

- Environment
  - constituents, impurities or additives in regulatory lists for PBT/vPvB concerns (identical or similar)
  - constituents, impurities or additives in external data sets for P, B and T properties (identical or similar)
  - partition coefficient data in registration dossiers
  - potential for bioaccumulation in terrestrial organisms (several scenarios using different predictive methodologies)
  - bioaccumulation data in registration dossiers
  - persistency data in registration dossiers
  - aquatic and terrestrial toxicity data in registration dossiers
  - QSAR models for P, B and T properties

Overview of hazard scenarios
Human health

- Human health
  - scenarios based on self classification for C, M, R, STOT RE and sensitisation
  - scenarios based on harmonised classification for C, M, R, STOT RE and sensitisation (identical or similar)
  - constituents, impurities or additives in regulatory lists for C, M, R, STOT RE and sensitisation (identical or similar)
  - constituents, impurities or additives in external lists indicating hazard, such as RoC, IARC, ISSCAN etc. (identical or similar)
  - structural alerts and QSAR models
  - hazard evidence in repeated dose, mutagenicity, carcinogenicity or reproductive toxicity studies in registration dossiers
  - NOAEL/LOAELs in registration dossiers

Hazard scenarios

Examples of external experimental datasets

	list	area / endpoint	author / source	common screening round created/updated
1	QSAR Toolbox BCF	Bioaccumulation	QSAR Toolbox	created in round 2
2	ISSCAN	Carcinogenicity	Italian national institute of health (ISS)	created in round 3
3	RoC	Carcinogenicity	U.S. Department of Health and Human Services	created in round 2
4	EDC SIN	Endocrine disruption	Chemsec	created in round 2
5	EDC COM	Endocrine disruption	European commission	created in round 2
6	EDC WHO	Endocrine disruption	World health organisation	created in round 2
7	EDC TedX	Endocrine disruption	TedX	updated in round 3
8	ToxCast ED	Endocrine disruption	U.S. Environmental Protection Agency	created in round 3
9	training sets of DK ED QSARs	Endocrine disruption	Danish Environmental Protection Agency	created in round 3
10	ISSCTA	Mutagenicity	Italian national institute of health (ISS)	created in round 3
11	ISSMIC	Mutagenicity	Italian national institute of health (ISS)	created in round 3
12	ISSSTY	Mutagenicity	Italian national institute of health (ISS)	created in round 3
13	QSAR Toolbox Kow	Partition coefficient	QSAR Toolbox	created in round 2
14	QSAR Toolbox P	Persistency	QSAR Toolbox	created in round 2
15	QSAR Toolbox Skin sens	Skin sensitisation	QSAR Toolbox	created in round 2
16	DART	Toxicity to reproduction	QSAR Toolbox	created in round 3
17	IMAP HH	C, M, R, STOT RE, skin sensitisation, respiratory sensitisation	Australian Department of Health	created in round 3
18	Canada challenge	Generic HH concern	Government of Canada	created in round 3
….				

the list will keep expanding subject to resource availability

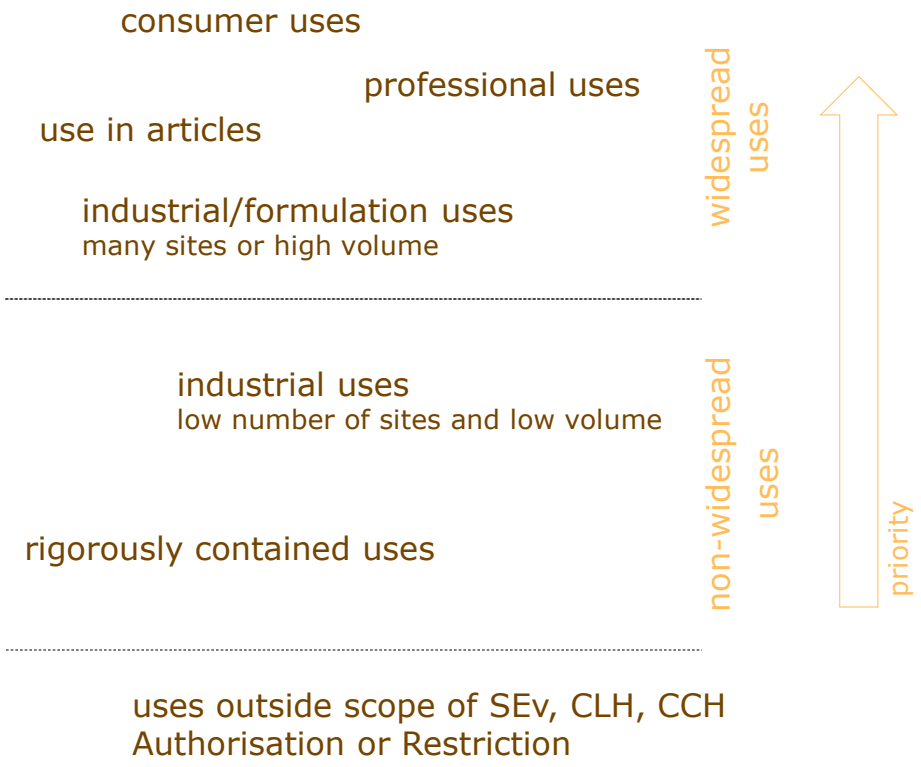
References

http://www.qsartoolbox.org/
http://www.iss.it/meca/index.php?lang=1&anno=2013&tipo=25
http://ntp.niehs.nih.gov/pubhealth/roc/roc13/index.html
http://chemsec.org/what-we-do/sin-list
http://ec.europa.eu/environment/endocrine/strategy/substances_en.htm#priority_list
http://www.who.int/ceh/publications/endocrine/en/
http://www.endocrinedisruption.org/
http://www.epa.gov/comptox/toxcast/data.html
http://qsar.food.dtu.dk/
http://www.chemicalsubstanceschimiques.gc.ca/challenge-defi/index-eng.php
http://www.nicnas.gov.au/chemical-information/imap-assessments/imap-assessments

Use and exposure

How do we prioritise hazardous substances for shortlisting?

- Identification of hazard is not sufficient for substance selection
- Use and exposure information is used to prioritise substances with high tonnage in wide dispersive uses



Guidance on information requirements and chemical safety assessment, chapter R.12: use description
http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf

Use and exposure

How can the existing approach be enhanced?

- The reliability of the approach can be further enhanced if we are able to automate the reliable retrieval of
  - the tonnage per use
  - the regulatory status of uses, such as uses in scientific research and development, plant protection products, biocidal products, fuels, cosmetics, food contact materials etc.
  - the number of sites where industrial and formulation uses take place
  - intermediate uses and strictly controlled conditions (outside Article 17 and 18 registrations)
  - estimated releases to the environment

- IUCLID 6.1 has been designed to enhance all these
  - the effectiveness of the prioritisation will also depend on the dossier updates we receive

Guidance on information requirements and chemical safety assessment, chapter R.12: use description
http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf

Common screening

General principles – substance grouping

- Common screening has invested significant resources in developing the necessary functionality for substance grouping
- Substance grouping is an essential element in the "Roadmap for SVHC identification and implementation of REACH risk management measures"[1] for both core and supplementary activities
- ECHA is using substance grouping as both evidence of hazard and as a tool to optimise regulatory output
- Substance grouping is based on structural similarity (using the generated molecular structures) and on proposed read-across and categories under REACH or other regulatory regimes

[1] SVHC Roadmap to 2020 Implementation Plan
http://echa.europa.eu/documents/10162/19126370/svhc_roadmap_implementation_plan_en.pdf

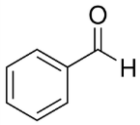
Structural clusters refined using structural alerts

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)

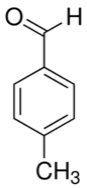
substance *i* (registered)



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

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

Substance grouping

Method 2: Use of read-across/categories

- Analogues 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

- All analogue information, regardless of its origin, was cast into the same format and added into the analogue library to facilitate both data collection and manual evaluation

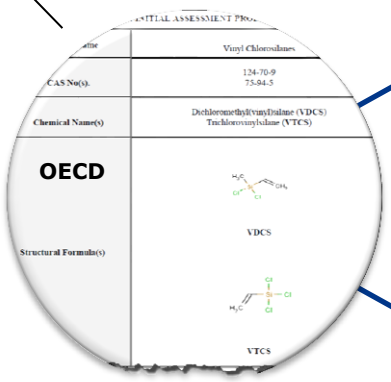
- The list of categories can be extended further
- How do you see the use of read/across and category information in screening?

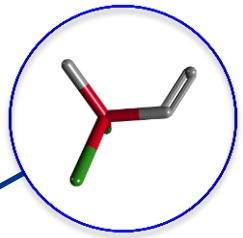
http://webnet.oecd.org/HPV/UI/ChemGroup.aspx
http://www.epa.gov/hpvis/
http://www.nicnas.gov.au/chemical-information/imap-assessments

Sources of analogue information
An example OECD category

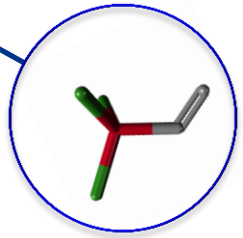
These two substances will be linked even if the registrations/notifications do not contain a read-across or category

vinyl chlorosilanes
(OECD category)



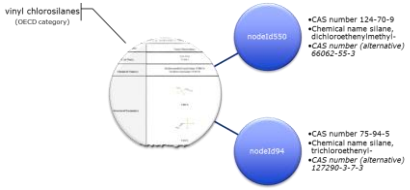


- CAS number 124-70-9
- Chemical name silane, dichloroethenylmethyl-
- *CAS number (alternative) 66062-55-3*



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

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



Using the analogue information

Linking registered/notified substances

for a given EC number we are interested in obtaining the following data:

- CAS number in REACH-IT
- CAS numbers in IUCLID
- alternative CAS numbers

for all registrations/notifications received

find all groups the substance belongs to

find all registered/notified substances belonging to these groups

we can control the sources of groups we want to utilise

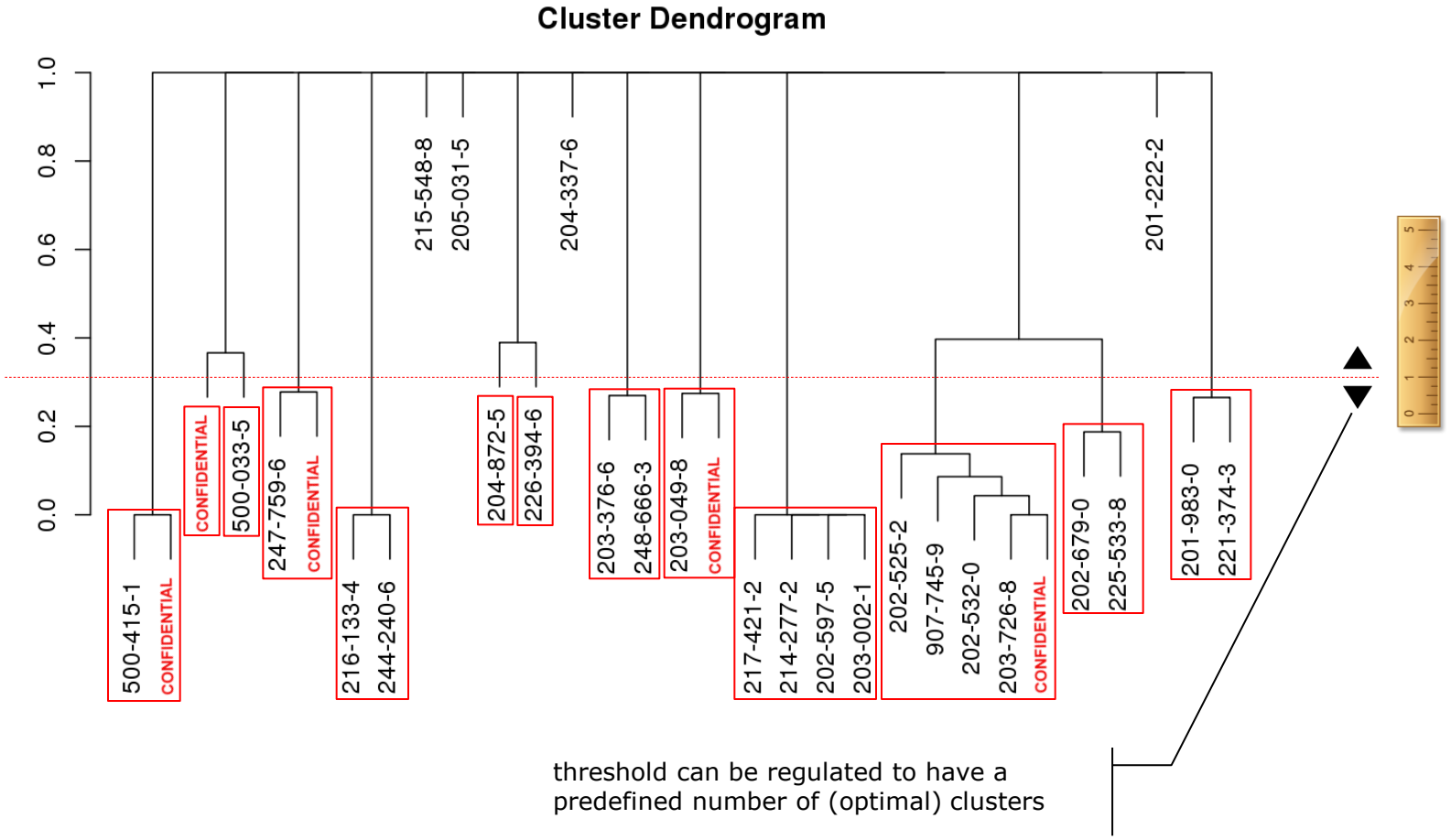
find all groups the analogues belong to

find all registered/notified substances belonging to these groups

we could extend the algorithm to provide n^{nt} generation analogues

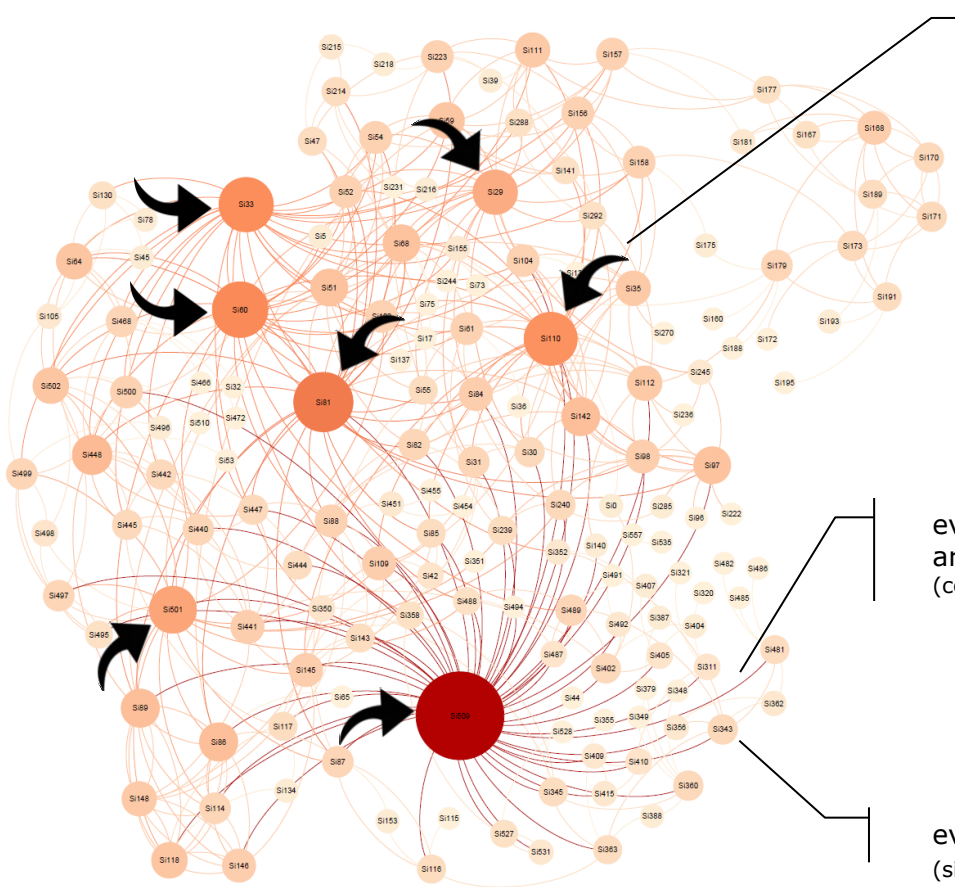
Substance grouping

Converting structural distances to dendrograms



Substance grouping

Visualising substance groups (silanes)



substances used most often as sources

every line is a read-across argument
(colour indicates the source substance)

every sphere is a substance
(size is proportional to the degree of linking)

Common screening

Documentation

- ECHA publishes a screening definition document[1]
  - hazard and use criteria
  - which external sources we use
- Updated annually, in consultation with Member States and industry stakeholders
- Contains the exclusion criteria per process and area (chapter 7)
- Contains the short listing criteria used to create the short list every year (chapter 8)
- Detailed technical description of the algorithms shared with Member States and EU COM

[1] Screening definition document (round 3)
http://echa.europa.eu/documents/10162/19126370/screening_definition_document_en.pdf

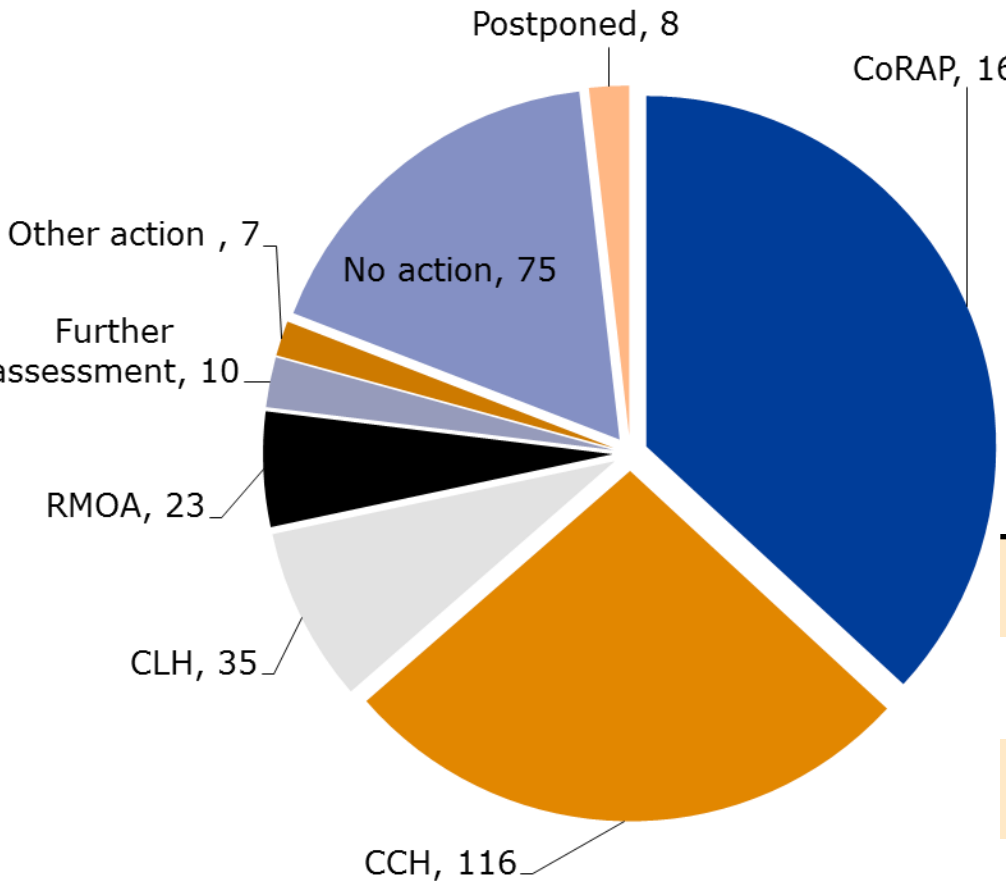
Content

- Background
  - definitions and timelines
- Tools and methods
  - underlying data used for screening
  - generation of molecular structures
  - overview of hazard scenarios
  - use and exposure
  - substance grouping
- Results obtained so far
  - outcome of rounds 2 and 3
  - progress with on-going round 3
- Conclusions

Several substances are pending because of a "parent" substance being under Substance Evaluation

Common screening

Results of manual screening, rounds 1 & 2



- 426 substances manually screened in rounds 1 & 2
- 82% found to require follow-up action
- majority require further information generation

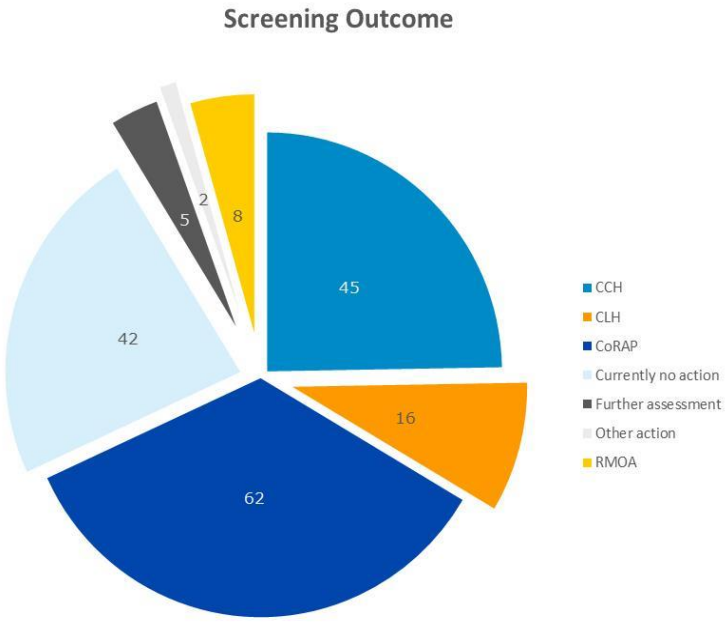
	CLH			
	pending action	intention	submitted dossier	public consultation
round 1	13	2	2	1
round 2	8	9	0	

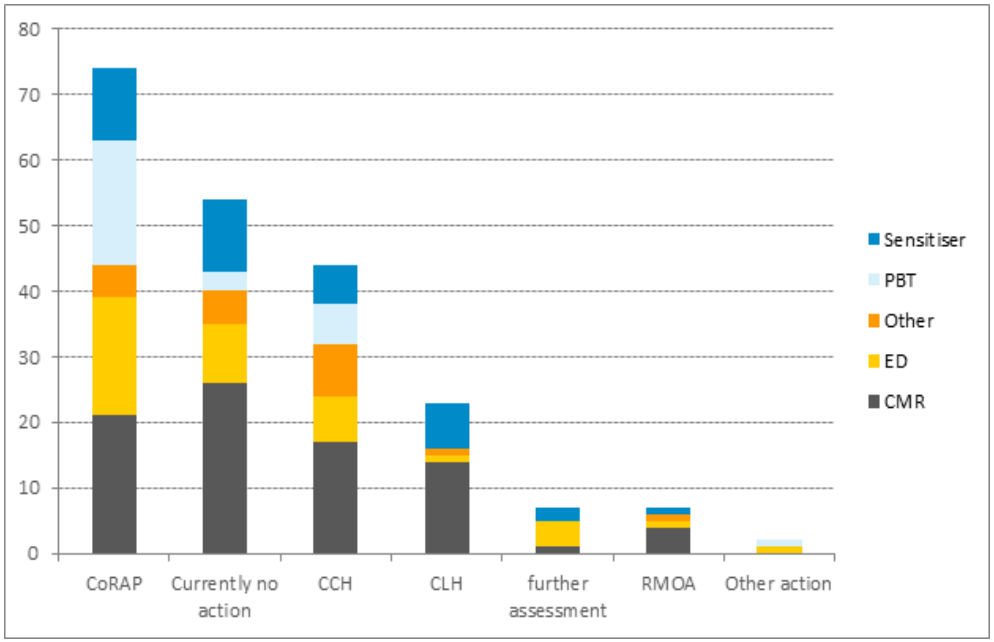
	RMOA			
	pending action	on going	SVHC intention	on candidate list
round 1	3	11	1	1
round 2	7	1	0	0

Common screening

Results of IT screening, round 3

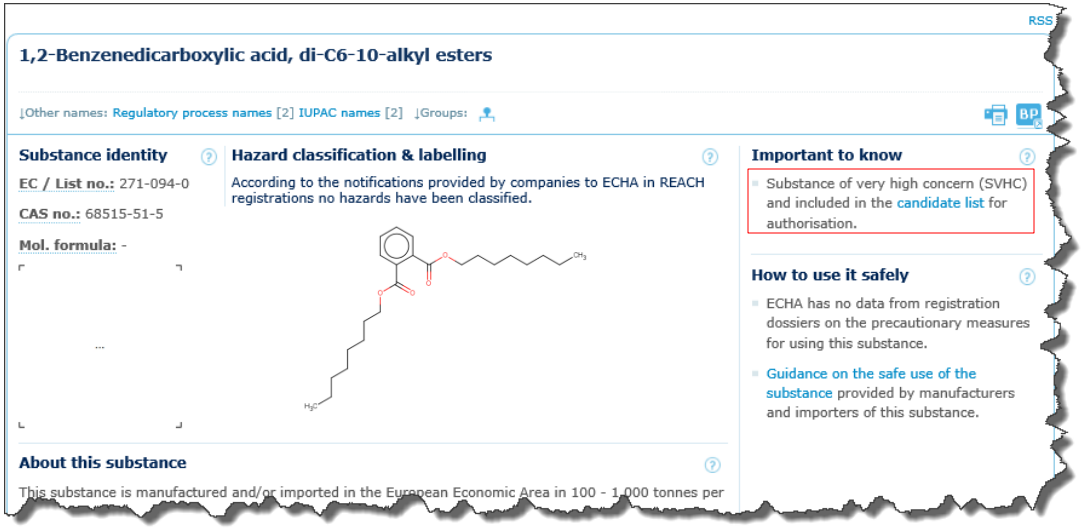
- 21 Member States participating
- manual screening on-going





Common screening

Example case I



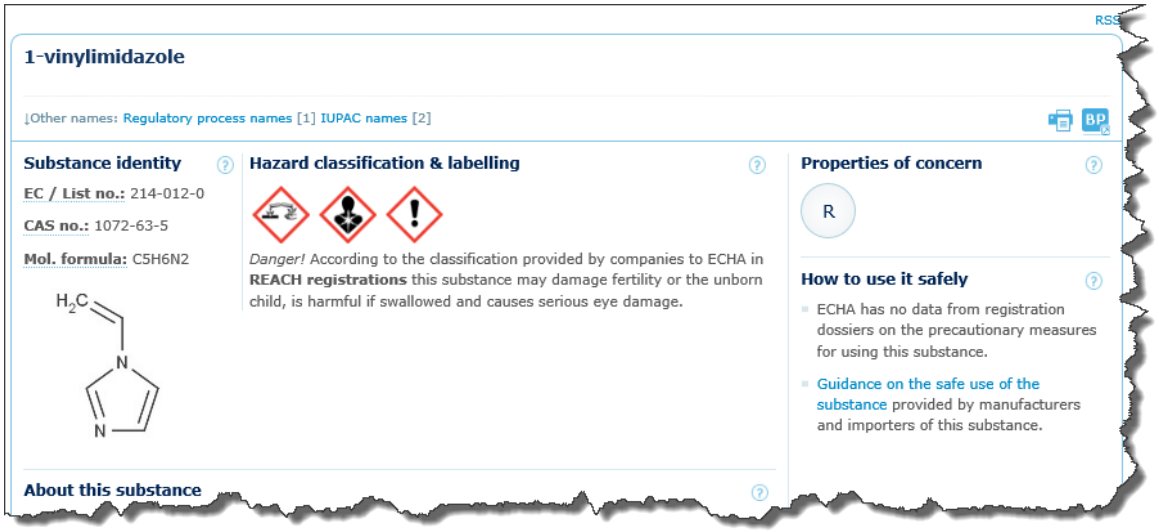
1,2-benzenedicarboxylic acid, di-C6-10-alkyl esters

- shortlisted in 2014 for CMR properties with indicative process as SEv
- manual verification concluded RMOA was needed
- RMOA completed in January 2015 concluded SVHC identification was needed based on the presence of dihexyl phthalate that is harmonised for reproductive toxicity
- substance added on the Candidate List in June 2015

ECHA
EUROPEAN CHEMICALS AGENCY

Common screening

Example case II



1-vinylimidazole

- shortlisted in 2014 for CMR with indicative process as SEv
- manual verification concluded no further information was needed and substance could be proposed for CLH
- CLH proposal submitted in Feb 2016 proposing classification as Repr. 1B – H360D
- public consultation on proposal ended on 4 April 2016

Conclusions

- Common screening is a flexible, agile cross directorate team with experts representing all REACH and CLP processes
- Common screening feeds practically all REACH and CLP processes with cases
  - the approach seems to focus on the right substances
- Comprehensive set of algorithms
  - deployed on a single platform
  - built to be modular and to enhance reusability
  - consultation with Member States and stakeholders
- Third year of operation; the process has been streamlined and understanding among ECHA and Member States is increasingly aligned
- Still, common screening is relatively new compared with the average time it takes to go through regulatory risk management or generate additional data