Substance identity - screening of the registration dossiers

Inconsistent identifiers

30 April 2014

Panos Karamertzanis
European Chemicals Agency

Substance identity IT screening campaign

Scope defined by 10 Issue types:

1. Missing concentration ranges
2. Typical concentration outside of concentration range
3. Composition reported with no constituents
4. Low or ambiguous degree of purity for well-defined substances
5. Unidentified constituent or impurity present at significant concentration
6. Well-defined substance with inconsistency between degree of purity and constituent concentrations
7. Well-defined substance with inconsistency between degree of purity and impurity concentrations
8. No spectral and analytical information provided
9. Additives without stabilising function
10. Inconsistent identifiers of constituents, impurities and additives

Why is this important?

- Chemical identifiers, such as IUPAC name and SMILES, are examined to establish the identity of the registered substance
  - when unclear ECHA may initiate a compliance check, or other regulatory processes such as testing proposal evaluation may be delayed

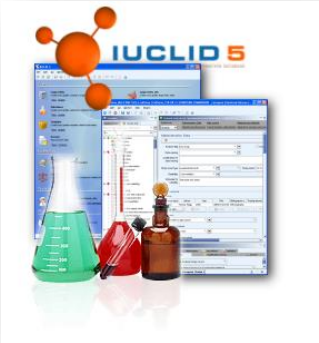
- Chemical identifiers are used by ECHA and Member State competent authorities to prioritise dossiers and substances for REACH and CLP processes
  - substance evaluation
  - risk management

Why is this important?

- Chemical identifiers are also used to (automatically) derive structures that are subsequently used in various ways
  - structural alerts to predict fate and (eco)toxicological properties
  - examine substance sameness in categories
  - examine structural similarity of the registered substance to substances of known concern

- For ECHA and Member State authorities, chemical identifiers in the registration dossier are much more than text strings
  - automatic conversion into structures
  - chemo-informatics
  - establish links with regulatory lists, such as Annex VI of the CLP Regulation (EC) No 1272/2008

Why is this important?



CAS number

108-88-3

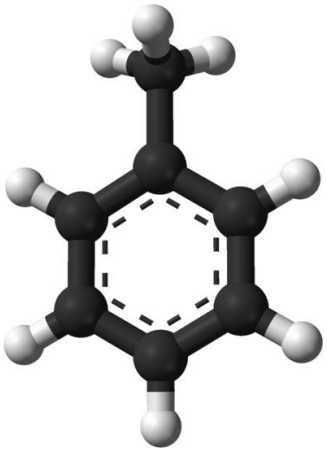
IUPAC name

methylbenzene

synonym

toluene

chemical structure used in screening for REACH and CLP processes

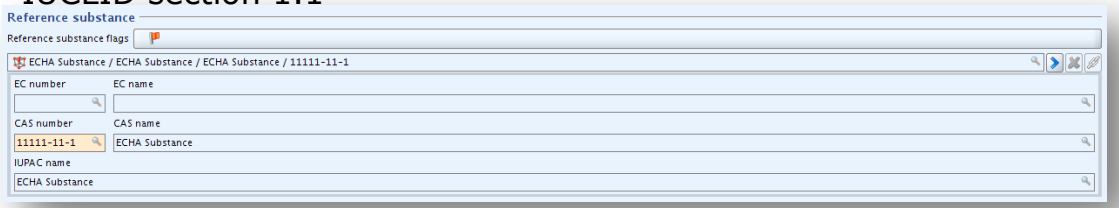


Which chemical identifiers are examined?

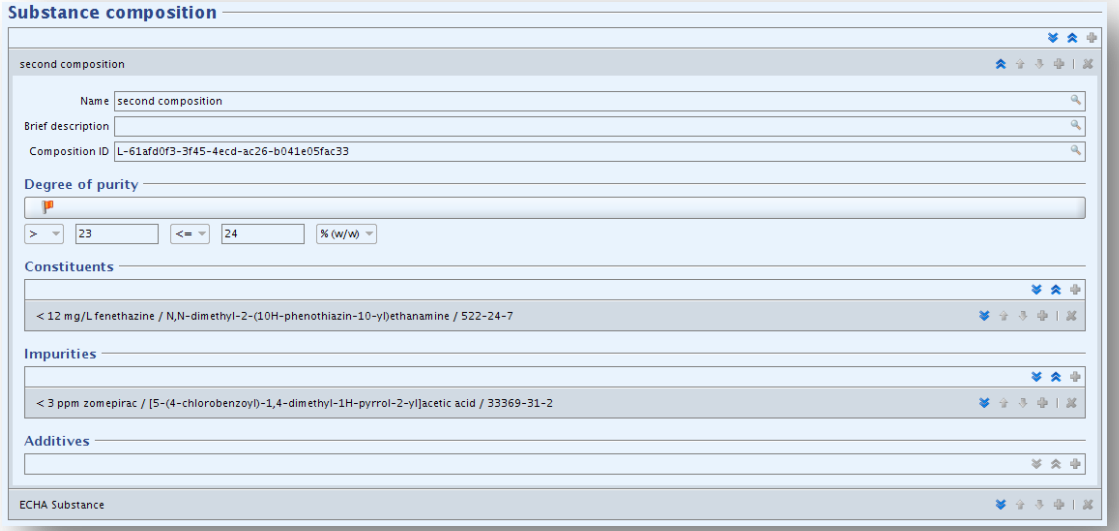
algorithms look in both the reference substance in section 1.1 and the individual constituents, impurities and additives in each composition in section 1.2

- 0 Related Information
- 1 General Information
  - **1.1 Identification**
  - 1.2 Composition
  - 1.3 Identifiers
  - 1.4 Analytical information
  - 1.5 Joint submission
  - 1.6 Sponsors
  - 1.7 Suppliers
  - 1.8 Recipients
  - 1.9 Product and process oriented research and development
- 2 Classification & Labelling and PBT assessment
- 3 Manufacture, use and exposure
- 4 Physical and chemical properties
- 5 Environmental fate and pathways
- 6 Ecotoxicological Information
- 7 Toxicological information
- 8 Analytical methods
- 9 Residues in food and feedingstuffs
- 10 Effectiveness against target organisms
- 11 Guidance on safe use
- 12 Literature search
- 13 Assessment Reports
- 14 Information requirements

IUCLID section 1.1

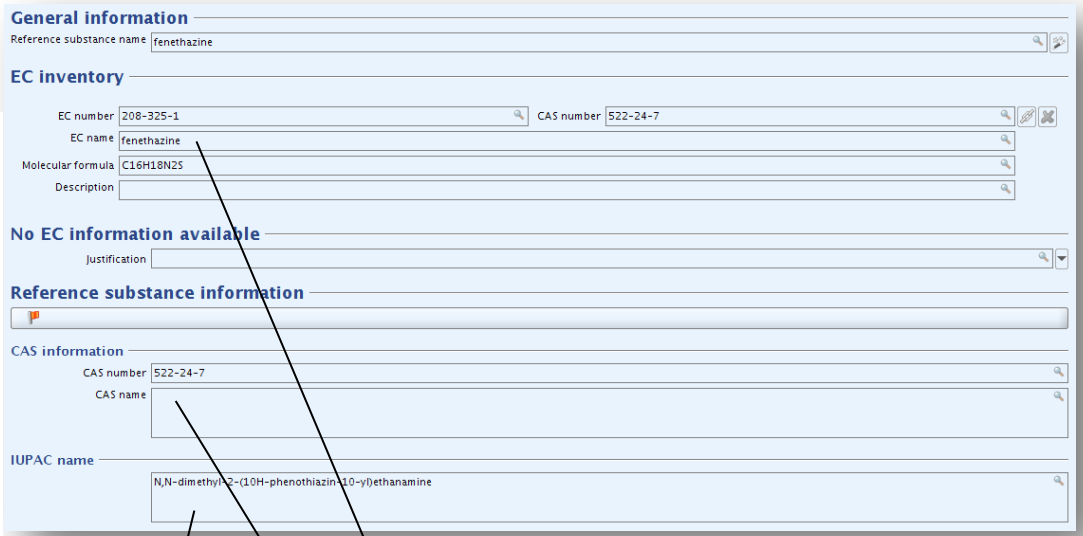


IUCLID section 1.2



Which chemical identifiers are examined?

- 0 Related Information
- 1 General Information
  - **1.1 Identification**
  - 1.2 Composition
  - 1.3 Identifiers
  - 1.4 Analytical information
  - 1.5 Joint submission
  - 1.6 Sponsors
  - 1.7 Suppliers
  - 1.8 Recipients
  - 1.9 Product and process oriented research and development
- 2 Classification & Labelling and PBT assessment
- 3 Manufacture, use and exposure
- 4 Physical and chemical properties
- 5 Environmental fate and pathways
- 6 Ecotoxicological Information
- 7 Toxicological information
- 8 Analytical methods
- 9 Residues in food and feedingstuffs
- 10 Effectiveness against target organisms
- 11 Guidance on safe use
- 12 Literature search
- 13 Assessment Reports
- 14 Information requirements



EC number and EC name

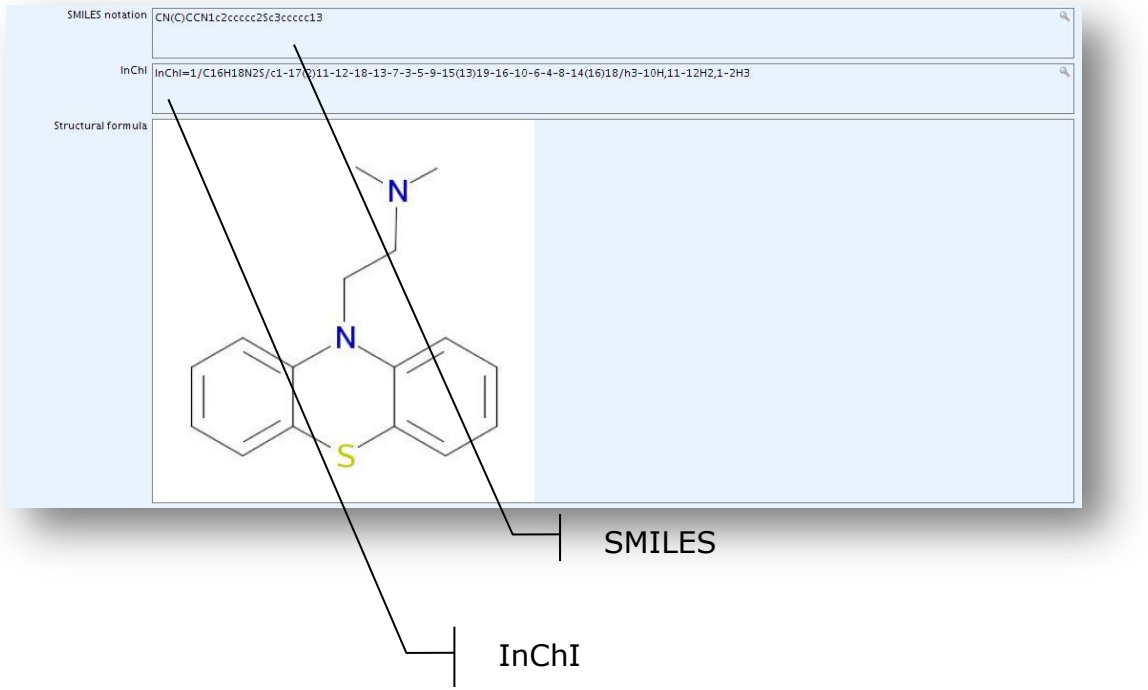
CAS number and CAS name

add here the correct CAS information if you disagree with the the EC inventory + explanation in the reference substance remarks

IUPAC name

Which chemical identifiers are examined?

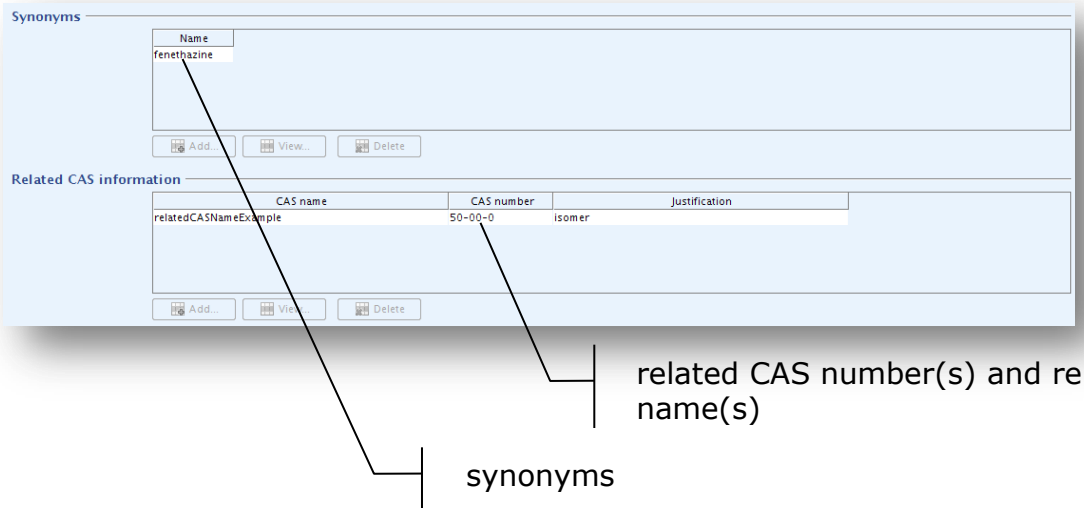
- 0 Related Information
- 1 General Information
  - **1.1 Identification**
  - 1.2 Composition
  - 1.3 Identifiers
  - 1.4 Analytical information
  - 1.5 Joint submission
  - 1.6 Sponsors
  - 1.7 Suppliers
  - 1.8 Recipients
  - 1.9 Product and process oriented research and development
- 2 Classification & Labelling and PBT assessment
- 3 Manufacture, use and exposure
- 4 Physical and chemical properties
- 5 Environmental fate and pathways
- 6 Ecotoxicological Information
- 7 Toxicological information
- 8 Analytical methods
- 9 Residues in food and feedingstuffs
- 10 Effectiveness against target organisms
- 11 Guidance on safe use
- 12 Literature search
- 13 Assessment Reports
- 14 Information requirements



please note that a valid InChI starts with 'InChI=1' or 'InChI=1S'

Which chemical identifiers are examined?

- 0 Related Information
- 1 General Information
  - **1.1 Identification**
  - 1.2 Composition
  - 1.3 Identifiers
  - 1.4 Analytical information
  - 1.5 Joint submission
  - 1.6 Sponsors
  - 1.7 Suppliers
  - 1.8 Recipients
  - 1.9 Product and process oriented research and development
- 2 Classification & Labelling and PBT assessment
- 3 Manufacture, use and exposure
- 4 Physical and chemical properties
- 5 Environmental fate and pathways
- 6 Ecotoxicological Information
- 7 Toxicological information
- 8 Analytical methods
- 9 Residues in food and feedingstuffs
- 10 Effectiveness against target organisms
- 11 Guidance on safe use
- 12 Literature search
- 13 Assessment Reports
- 14 Information requirements



Which chemical identifiers are examined?

structure origin and conversion method	number of successful conversions
I5_CAS (I)	29566
I5_CAS (II)	106203
I5_CASNAME (I)	32217
I5_CASNAME (III)	35469
I5_EC (I)	28011
I5_ECNAME (I)	86405
I5_ECNAME (III)	94820
I5_ECNAME (IV)	118633
INCHI (I)	80762
INCHI (III)	76529
IUPACNAME (I)	84900
IUPACNAME (III)	87134
IUPACNAME (IV)	87323
REF_SUB_NAME (I)	89312
REF_SUB_NAME (III)	97986
REF_SUB_NAME (IV)	91449
SMILES (I)	94842
SMILES (III)	95670

numbers are indicative

chemical structures

...

Erroneous identifiers, such as invalid IUPAC names, or inconsistent identifiers cause issues and can be one of the reasons to manually examine the dossier in greater detail

Messages in information letters

Excerpt from the information letter:

"Inconsistent identifiers of constituents, impurities and additives

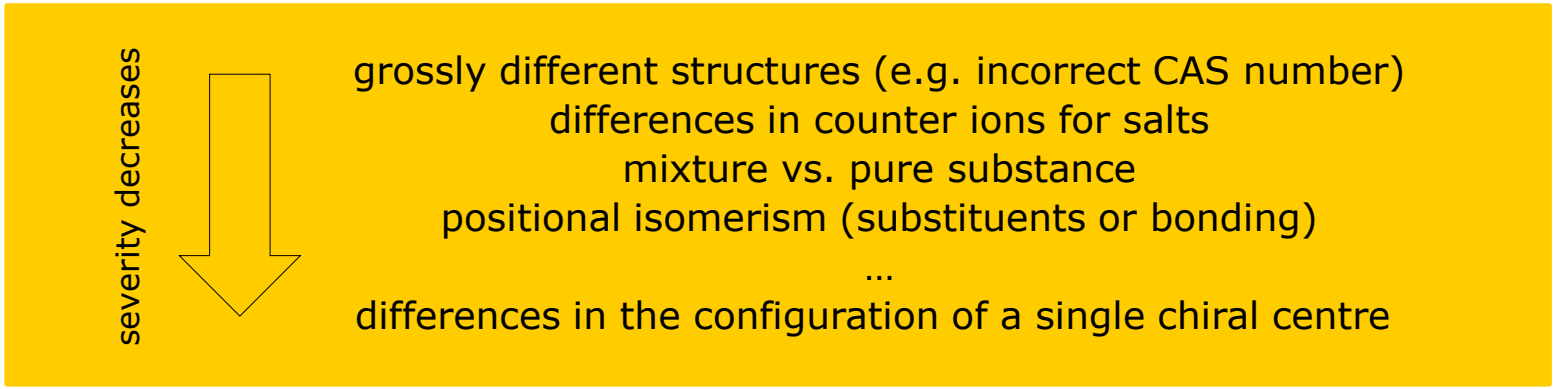
The impurity with reference substance name *"Reference Substance Name"* and IUPAC name *"IUPAC Name"* empty in the composition with name *"Reference Substance Name"* and local UUID *"Lf5cecc8b-2bce-4f0d-a552-45801ccc6785"* seems to have two sets of inconsistent identifiers leading to different molecular structures:

- molecular structure 1 is derived from the identifier SMILES *"CCCCCO …CCCC(=O)O"*
- molecular structure 2 is derived from the following two identifiers CAS number *"123-123-12"* and InChI *"InChI=1/C17H34O2/c1-2-… (H,18,19)"*

two groups of chemical identifiers leading to different molecular structures

Automatic identification of inconsistent identifiers

- All reference substances, in all compositions are checked separately
- The chemical identifiers are processed and used to generate a molecular structure
  - 'textual' identifiers, such as IUPAC names are also used and automatically converted to structures
  - 'numerical' identifiers, such as CAS numbers are also used to generate a molecular structure using reliable, external sources
  - SMILES and InChIs are directly used as they provide structural information
- The molecular structures produced by the different identifiers are checked for consistency
  - inconsistencies are grouped in terms of severity
  - type of inconsistency, and
  - number of distinct molecular structures derived

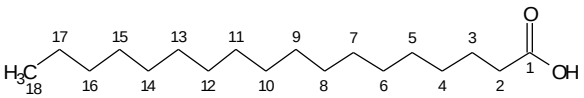


Key messages I

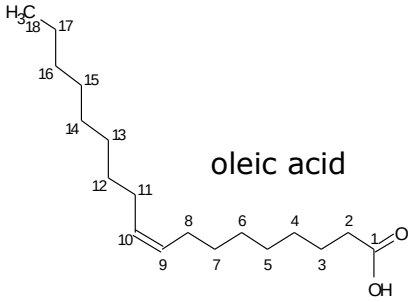
- Check the identifiers for all reference substances in all compositions in your registration dossier, regardless of their concentration
  - even if you have not received an information letter
- Do not use generic identifiers (e.g. EC number or CAS number), if your reference substance can be identified with a more specific identifier
- On some occasions, the identifiers in the EC inventory or prefilled reference substances may not be perfectly accurate; it is your responsibility to check their correctness
  - IUCLID allows a different CAS number than the one in the EC inventory to be specified
  - IUCLID allows the imported reference substance to be edited, e.g. to change the SMILES
- **Please pay attention to all aspects**
  - enantiomers (R/S), or other isomers (e.g. cis/trans), are not the same structure, although a CAS number does not necessarily specify the isomerism
  - for further information, please consult the Guidance on substance identification1 (Chapter 5) and the Dossier Submission Manual, part 182 (DSM 18, Q&A 4, Q&A 7)
  - SMILES and InChI discriminate among isomers if used properly

Key messages II

- Differences in branching, or differences in the position of an unsaturated bond, mean different substances, e.g.
  - EC number 200-313-4, stearic acid ($C_{18}H_{36}O_2$) ≠ EC number 204-007-1, oleic acid ($C_{18}H_{34}O_2$)

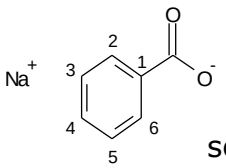


stearic acid

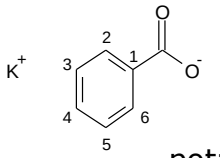


oleic acid

- A neutral acid or base is not the same with an individual salt
- Salts with different counterions are different substances, e.g.
  - EC number 208-534-8, sodium benzoate ($C_7H_5O_2.Na$) ≠ EC number 209-481-3, potassium benzoate ($C_7H_5O_2.Ka$)



sodium benzoate



potassium benzoate

[1] Data Submission Manual, part 18, how to report the substance identity in IUCLID 5 for registration under REACH, echa.europa.eu/documents/10162/13653/substance_id_report_iuclid_en.pdf

[2] Guidance for identification and naming of substances under REACH and CLP, echa.europa.eu/documents/10162/13643/substance_id_en.pdf

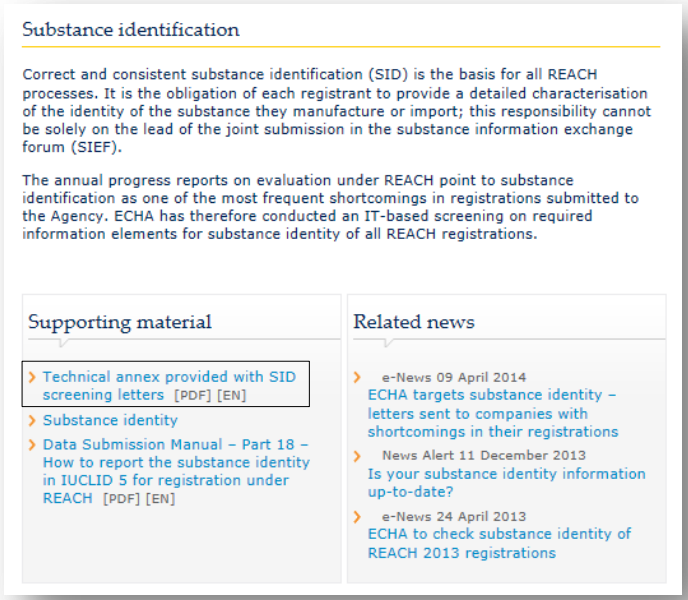
References

- **Guidance for identification and naming of substances under REACH and CLP**
  echa.europa.eu/documents/10162/13643/substance_id_en.pdf

- **Data Submission Manual, part 18, How to report the substance identity in IUCLID 5 for registration under REACH**
  echa.europa.eu/documents/10162/13653/substance_id_report_iuclid_en.pdf

- **Questions and answers on substance identification**
  echa.europa.eu/qa-display/-/qadisplay/5s1R/view/REACH/SubstanceIdentification

- **How to improve your dossier webpages**
  echa.europa.eu/support/how-to-improve-your-dossier



Thank you

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

Follow us on Twitter
@EU_ECHA

Follow us on Facebook
Facebook.com/EUECHA