1. Data curation

Manually curated

Chemistry olympiads

University exams

University exercise sheets

Manually created

Semi-programatically curated

GHS pictograms

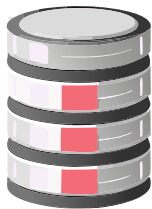Daily allowed intakes

Hazard statements

Number of NMR peaks

Electron counts

IUPAC-SMILES questions

Oxidation states

Point groups

2. Semantic annotation

Reactions

`[START_RXNSMILES] [END_RXNSMILES]`

Molecules

`[START_SMILES] [END_SMILES]`

`\ce{C6H6}`

Units

`\pu{m^{-3}}`

Equations

`$a^2 + b^2 = c^2$`

3. Review

Manual inspection

Factual correctness

Clarity and phrasing

Error analysis

Automatic checks

Schema

Invariance to shuffling

Spelling