

## 1. Data curation

### Manually curated

Chemistry olympiads

University exams

University exercise sheets



### Semi-programmatically 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

$\mu\text{m}^{-3}$



### 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

