# Reaction Data Curation I: Chemical Structures and Transformations Standardization

* Abstract: - the quality of experimental data for chemical reactions is a critical, in this the researchers suggest a 4 steps protocol that includes the curation of individual structures, chemical transformations, reaction conditions, and endpoints.
* Introduction: - Data quality is crucial for the effective storage and exploitation of chemical knowledge.
* So far, two databases have been intensively used in the modelling studies: Reaxys and USPTO reaction dataset.
* The Reaxys dataset is a commercial data set, it has been extensively used for various applications of deep-learning neural networks to retrosynthesis, robochemistry, and prediction of optimal reaction conditions, as well as for analysis of rection network.
* The USPTO is the largest public dataset of chemical reactions extracted from the US patents using text mining techniques.
* The dataset has been extensively used for different applications: analysis of reaction databases, forward and backward-synthesis, reactions classification, atom-to-atom mapping, yield prediction, and compound role assessment.
* Very little effort has been invested to curate reaction data.
* Most of the suggested structures’ standardization workflows, recombine selected steps of the molecules cleaning proposed suggested by Fourches et al.



* A *transformation* term corresponds to a reaction equation that links sets of reactants and products.
* According to IUPAC definition, a *reactant* is “a substance that is consumed in the process of a chemical reaction.”
* Whereas the *product* is “a substance that is formed during the chemical reaction”.
* Atom of reactants that change their connectivity and/or formal charge(s) upon the transformation constitute a *reaction center.*
* To identify a reaction center, a one-to-one correspondence between atoms of reactants and products called *atom-to-atom mapping* (AAM) must be established.
* Reactants are the catalyst, solvents, catalytic, poisons, complexation, agents, redox, agents, detergents, and acid/bases.
* According to IUPAC a reagent is a “substance that is added to the chemical system in order to bring about a reaction” is not precise enough for automatized reagents recognition.
* Schneider et al. attributed a compound to reagents if it wasn’t affected in the course of reactions, whereas Gao et al. considered any compound as a reagent if it did not contribute any carbon atom to the reaction.
* Normally a conditions’ description includes both numeric properties like temperature, pressure, and reaction duration, as well as reagents.