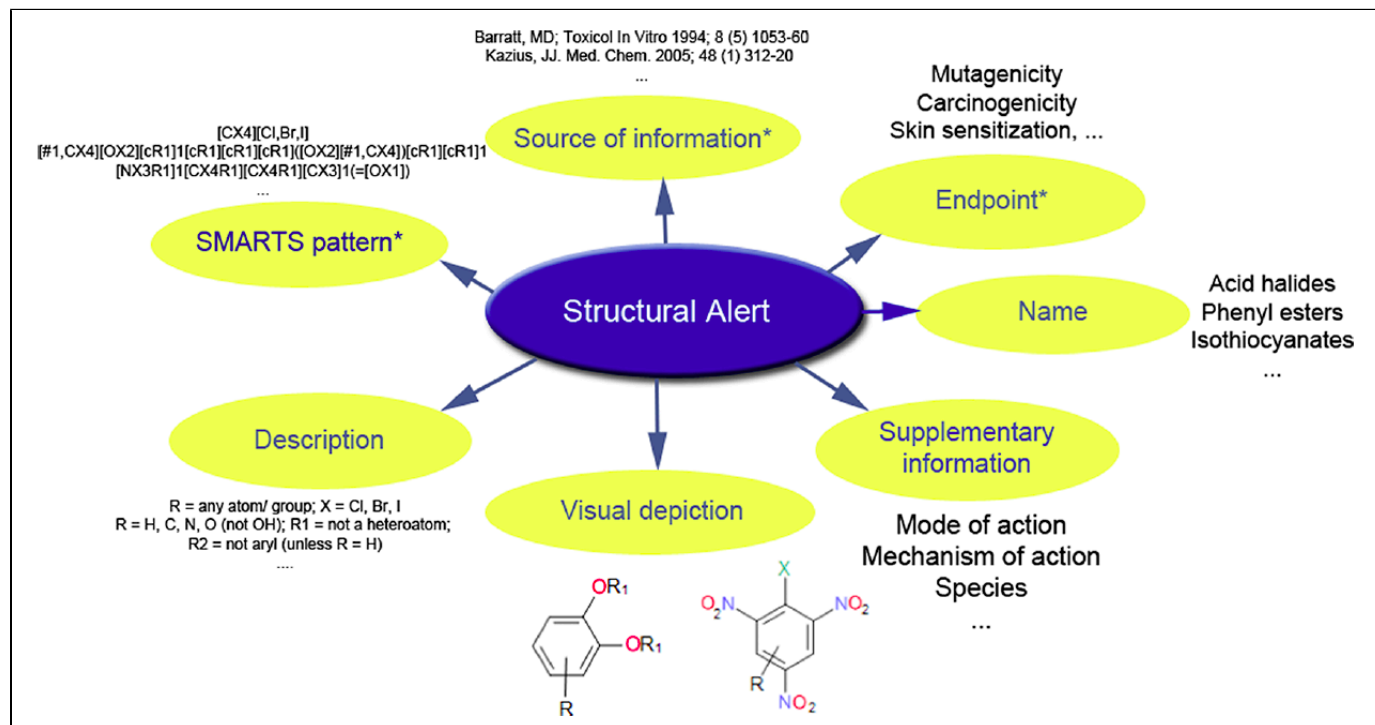# Definition of an alert

The central entity in the ToxAlerts database is a structural alert (a substructure or, more universally, a SMARTS string), which is uniquely identified by the following:

1. A structural pattern represented by a SMiles ARbitrary Target Specification (SMARTS) string.
2. A publication, where the alert was mentioned.
3. A toxicological endpoint associated with this alert (e.g.,carcinogenicity or skin sensitization).

In addition to these key components, the database is designed to store any other additional information, such as:

1. A chemical name of the compound class represented by the alert (e.g., " Acid halides" , " Sulphonyl azides" , etc.).
2. A visual depiction of the alert: since an automatic generation of a depiction from a pattern is an ambiguous and a nontrivial task, users can generate and upload their depictions in PNG format manually.
3. Position of the alert in the publication (i.e., page, table, line).
4. Arbitrary supplementary information (e.g., mechanism of action associated with the alert, species, metabolic activation information, etc.).

A simplified schema of the database is presented on the following figure:

SMARTS as alert patterns

As it was mentioned above, ToxAlerts uses SMARTS patterns to represent toxicological alerts. The major advantage of using SMARTS is that a compound can be matched against an alert in an automatic manner using one of the available chemical software libraries. For our purposes, we use the MolSearch utility from Chemaxon (www.chemaxon.com). Although SMARTS are not always easily interpretable by human, a visual depiction, the chemical name and the description provide a sound way to understand the alert.

Extended pattern syntax

In some cases, SMARTS patterns may not be sufficient to represent a desired alert. ToxAlerts provides and extended syntax, which supports logical operations (conjunction and inversion), restrictions on molecular weight and on count of molecular fragments. In brief, a user can write expression like

> (i) MW > 100 *AND* [NX3](=O)=O > 2

to match molecules with molecular weight of more than 100 and containing more than two nitro groups.