ToxML: a structured and serialisable exchange standard for toxicology

#861

Mohammed A Ali[1], Mukesh L Patel[1], Philip N Judson[2], R Daniel Benz[3]

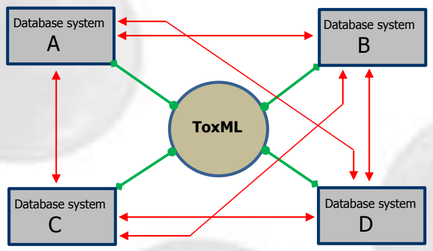
[1]Lhasa Limited, Leeds, UK, [2]Judson Consulting Service, Harrogate, UK;, [3]FDA Center for Drug Evaluation and Research, Silver Spring, MD USA

[The views expressed here are those of the authors; this is not an official US FDA guidance or policy statement]

Abstract

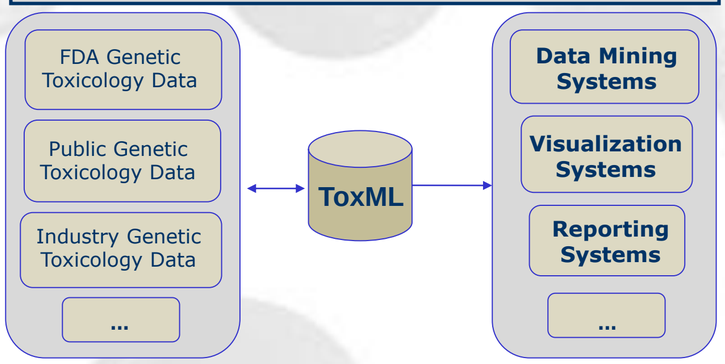
The amount of toxicity data available for those wishing to share and communicate knowledge, or to use for data mining and modelling, is continually growing within the biomedical disciplines. The challenge with this expanding amount of data is that it exists in a multitude of different formats. ToxML is an open standard based on Extensible Markup Language (XML) that consists of an XML Schema (XSD) defining the toxicology schema and lists of controlled vocabulary that ensure consistency of usage. The use of XML means that the data can be created, stored and transported in a structured format that is not bound to a specific software application or programming language. The data file model resulting from this approach is very versatile and allows for the aggregation of experimental data up to the compound level in the detail needed to support areas such as quantitative structure-activity relationship (QSAR) development. ToxML formats have been developed, so far, for 27 toxicity study types. These cover both *in vivo* and *in vitro*, and currently include the following super toxicity endpoints: genetic toxicity, carcinogenicity, skin sensitisation, skin penetration, *in vivo* repeat dose toxicity, *in vivo* single dose toxicity and ecotoxicity. Informatics groups at the FDA CDER-CFSAN use ToxML to populate repositories with the results of FDA toxicological and clinical data harvesting efforts, and employ the resulting information to model QSARs. We describe an example of how ToxML can be used as a practical data exchange standard for genetic toxicology information. The standard is maintained by a curation team overseen by the ToxML organisation. The standard is published on a web site (www.toxml.org) together with tools to allow members of the user community to view and download it and to contribute to the next version. An open forum is also provided via a wiki on the website.

Purpose of Standard



A common exchange standard means that only one data exchange interface module is needed for each application, and therefore removes the need for multiple conversions between database systems (red arrows).

Data ⟶ Integration ⟶ Analysis



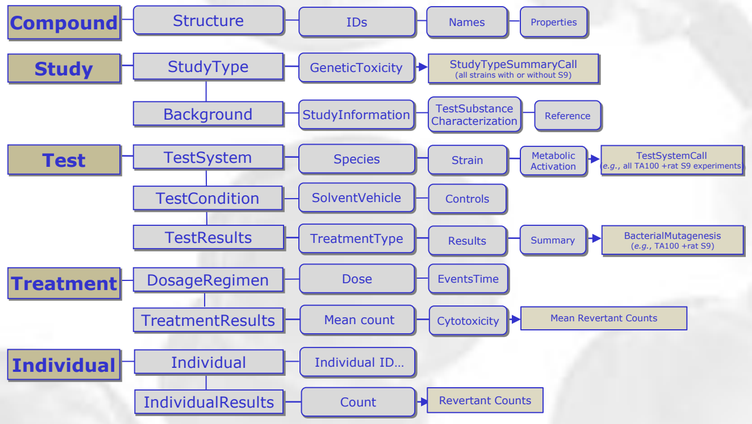
Acknowledgements

- FDA Center for Food Safety and Applied Nutrition (CFSAN)
- Leadscope *In Silico* Toxicology (LIST) consortium
- National Institute of Standards and Technology / Advanced Technology Program (NIST/ATP)
- OpenTox (European interoperable predictive toxicology framework)

ToxML Characteristics

- Design is open standard
- Covers single and repeat dose genetic toxicity and other toxicity studies
- Data format based on extensible markup language (XML)
  - Hierarchical data structure which lends itself well to modeling data of this nature
  - Chemical substance serves as the fundamental base key, or root node
  - Supports both binary & non-binary data (numbers, text, images)
- Toxicity studies are submitted as self-contained data files for one or more studies
  - Record keys are predefined due to the hierarchical structure
- Uses controlled / normalized vocabulary to ensure harmonization

Bacterial Mutagenesis ToxML Structure



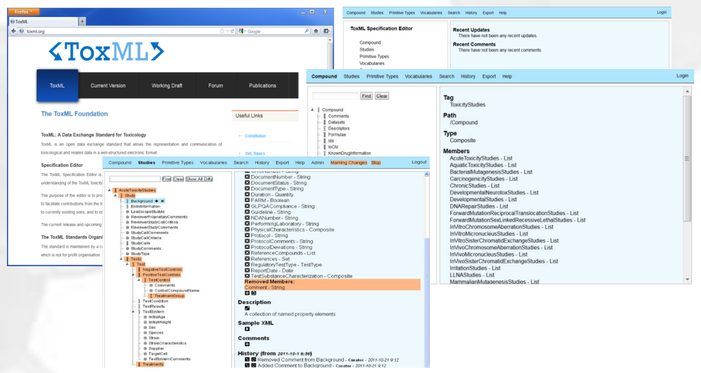
***Toxicology Mechanisms and Methods* 18:103–118, 2008**

ToxML Representation (Sample)

```
<Compound version="3.0.17">
 <Ids>
  <Id type="cas">101904-52-3</Id>
 </Ids>
 <Names>
  <Name sources="cfsan-ofas" type="chemName">2-Methyl-7-toluidinoquinacridone-6-carboxylic acid</Name>
  <Name sources="cfsan-ofas" type="synonym">TMAS;  3-Acridinecarboxylic acid, 9,10-dihydro-7-methyl-2-[(4-methylphenyl)amino]-9-oxo</Name>
 </Names>
 <ToxicityStudies>
<BacterialMutagenesisStudies>
   <Study sources="cfsan-ofas">
    <StudyType>bacterial mutagenesis</StudyType>
    <StudyCalls>
     <Call>Negative</Call>
    </StudyCalls>
    <Background>
     <Confidentiality>No</Confidentiality>
     <PhysicalCharacteristics>
      <TestMatrix>Neat</TestMatrix>
      <TestForm>Parent</TestForm>
      <PhysicalState>Powder</PhysicalState>
      <StorageCondition>dark at 22 Celsius</StorageCondition>
     </PhysicalCharacteristics>
    </Background>
    <Tests>
     <Test>
      <TestSystem>
       <Species>Salmonella typhimurium</Species>
       <Strain>TA1538</Strain>
       <MetabolicActivation>Absent</MetabolicActivation>
      </TestSystem>
      <TestCondition>
       <SolventVehicle>
        <Substance>
         <Substance>Dimethylsulfoxide (DMSO)</Substance>
        </Substance>
       </SolventVehicle>
       <Solubility>>2500 ug/plate</Solubility>
       <AssayTechnique>plate incorp.</AssayTechnique>
      </TestCondition>
      ...
```

```
(cont.)
...
<Treatments>
 <TreatmentGroup>
  <DosageRegimen>
   <Dose>
    <Value>4.0</Value>
    <Units>micro-g/plate</Units>
   </Dose>
   <TreatmentTime>
    <Value>
     <LowValue>48.0</LowValue>
     <HighValue>72.0</HighValue>
    </Value>
    <Units>hour</Units>
   </TreatmentTime>
   <DoseDuration>
    <Value>
     <LowValue>48.0</LowValue>
     <HighValue>72.0</HighValue>
    </Value>
    <Units>hour</Units>
   </DoseDuration>
  </DosageRegimen>
  <IndividualSubjectCount>20</IndividualSubjectCount>
  <Results>
   <Precipitation>No</Precipitation>
   <Cytotoxicity>No</Cytotoxicity>
   <RevertantCount>83</RevertantCount>
   <RevertantCountSignificance>4.2</RevertantCountSignificance>
  </Results>
 </TreatmentGroup>
 ...
```

Specification Viewer & Editor



The specification is freely available and allows registered parties to make contributions. The forum section of the website facilitates the open discussion and evolution of the specification.

Industry and Regulatory Agency Needs for ToxML

- Data mining software and scripts require well-structured electronic data
- Data visualisation requires well-structured electronic data
- *In silico* predictive toxicology requires well-structured electronic data

Current Genetic Toxicology ToxML Endpoints

- Bacterial mutagenesis
- *In vitro* chromosome aberration
- *In vitro* micronucleus
- Mammalian mutagenesis
- *In vivo* chromosome aberration
- *In vivo* micronucleus

Users of ToxML with Genetic Toxicity Information

- US FDA / Center for Food Safety and Applied Nutrition
  - Internal data exchange and archiving
  - Data export for public use
- US FDA / Center for Drug Evaluation and Research
  - Internal data exchange and archiving
  - Data export for public use
- Lhasa Limited
  - Integrate US FDA data into VITIC
- Boehringer Ingelheim
  - Internal data exchange and archiving
- Named as preferred data exchange format in eTOX project plan
  - Drug safety database being developed to support the EU Innovative Medicines Initiative for the pharmaceutical industry

Special ToxML Features

- Enables report generation that can create human-readable toxicity study reports from the underlying ToxML data records
- Allows integrated and normalised data mining both for searching and algorithmic endpoint calculation
- Can include binary data directly within the XML of ToxML so that images (*e.g.* histopathology slides) can be included within the context of a particular record instead of relying on a loosely coupled file system approach used in other standards (e.g. SEND)

Future Plans for ToxML

Lhasa Limited has administrative duties for ToxML

- ToxML fields and vocabulary were originally chosen by the Leadscope *In Silico* Toxicology consortium.
- ToxML is now maintained by the not-for-profit ToxML Standards Organisation (TSO). Lhasa Limited are currently managing work on the project on behalf of the TSO, which is overseen by an Advisory Board.

Meeting the needs of our users

- The ToxML Standards Organisation and the ToxML website are designed to ensure that the development of the standard is driven by the users.
- Please make the additions and changes that best suit you, and influence discussions and policy decisions, by visiting the website, http://www.toxml.org, where you can use the editor, contribute to the forum, and email your views to us.

Add new endpoints

- This can be facilitated by the users; allowing the toxicology community to develop the standard incrementally, by concentrating on those areas where it is needed and thus progressively building a broader standard.

References:

Ali MA, et al. *ToxML, a data exchange standard with content controlled vocabulary used to build better (Q)SAR models*, SAR QSAR Environ Res (in press)