ToxML: A Common Exchange Standard for Raw and Summarized Genetic Toxicology Data and Data Analysis Results

P49

R. Daniel Benz[1], Kevin P. Cross[2], David A. Bower[2], Philip N. Judson[3], Christopher G. Barber[4], Shree Nath[5]

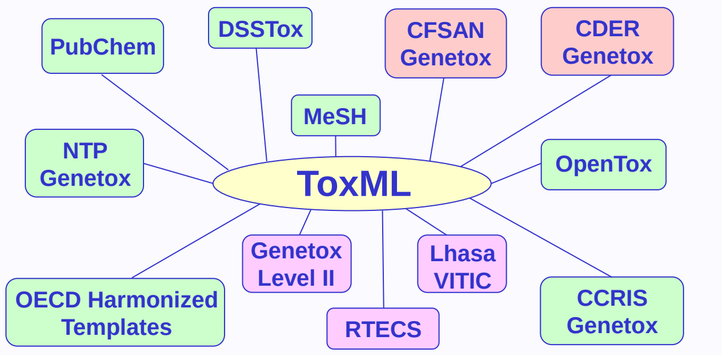
[1]*FDA Center for Drug Evaluation and Research, Silver Spring, MD USA;* [2]*Leadscope, Inc.,, Columbus, OH USA*; [3]Judson Consulting Service, Harrogate UK; [4]Lhasa Limited, Leeds UK; [5]PointCross Life Sciences, Foster City, CA USA

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

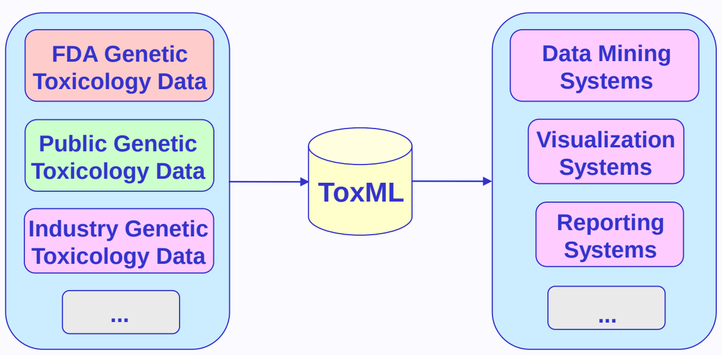
Abstract

The volume and complexity of toxicity datasets that scientists wish to share or use in data mining, analytics and modeling is continually increasing. The challenge with exploiting this valuable resource is that this information is currently locked in a multitude of proprietary formats, depending on the source and original purpose. Different data capture systems covering the same endpoint often use different conventions to structure and represent the same information. Restructuring datasets to effectively analyze, combine and share them is laborious, time consuming and error-prone. The ToxML project addresses the need for an open source data exchange standard that represents toxicological data in a structured electronic format. ToxML is a well-defined, XML-based data standard featuring consistent naming conventions and controlled vocabularies. The model is versatile and extensible, allowing the representation of study data to suit multiple end uses. ToxML provides a consistent, readily transferable representation of compound data, protocols, raw data, summary data, and data analysis results. ToxML files can be used by any application capable of incorporating XML information. A toolkit for parsing ToxML is publicly available. The USFDA Center for Food Safety and Applied Nutrition and Center for Drug Evaluation and Research use ToxML to populate repositories with the results of FDA toxicological and clinical data harvesting efforts, and employ the resulting information to model (quantitative) structure-activity relationships. This poster shows examples of how ToxML is being and could be used as a practical data exchange standard for genetic toxicology information.

Data Integration (Examples)



Data → Integration → Analysis



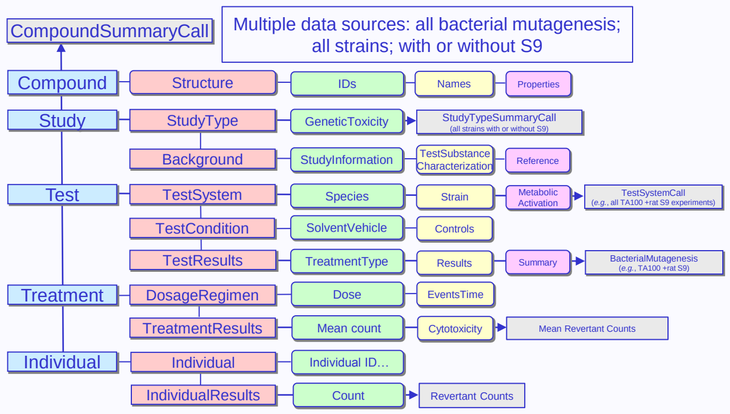
Acknowledgements

- 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 quite well to modeling data of this nature
  - Chemical substance serves as the fundamental base key, or root node
  - Supports binary data which could be used to include items like 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



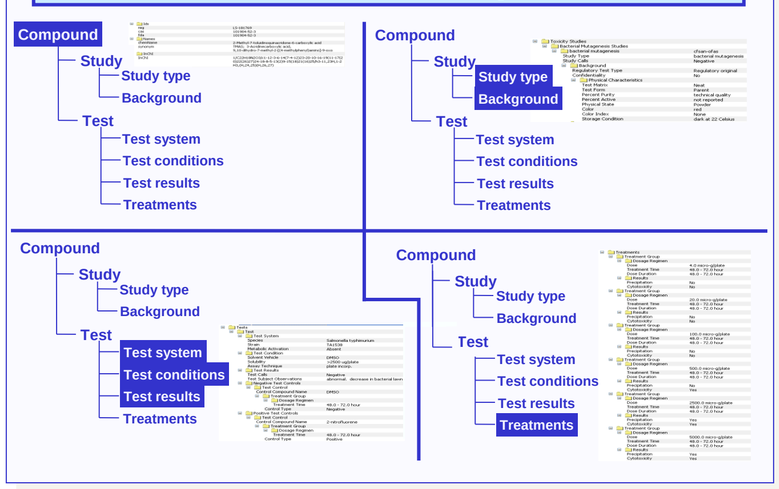
Toxicology Mechanisms and Methods **18:103–118, 2008**

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>
  …
```

ToxML Record Examples



Industry and Regulatory Agency Needs for ToxML

- Data mining software and scripts require well-structured electronic data
- Data visualization 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
- Recommended for use by eTOX
  - 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 normalized 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

Future Plans for ToxML

- Lhasa Limited has assumed administrative duties for ToxML
  - ToxML fields and vocabulary were originally chosen by the Leadscope *In Silico* Toxicology consortium. Lhasa is now establishing a process and mechanism through which ToxML can be expanded by reviewing current needs and input from contributors (http://www.toxml.org).
- Ensure a loss-less compatibility in converting other formats to ToxML
  - Identify additional data fields as needed
  - Harmonize controlled vocabularies
  - Ensure compatibility with the FDA Standards of Exchange of Non-clinical Data (SEND) format by harmonizing vocabulary terms and findings, and adding individual animal data to the ToxML test level information
  - Develop compatibility with the OECD harmonized templates
- Develop ToxMLs for additional endpoints
  - A ToxML format for *in vitro*, *in vivo*, and clinical cardiovascular effects is being created
  - A ToxML for 'omics data is being considered