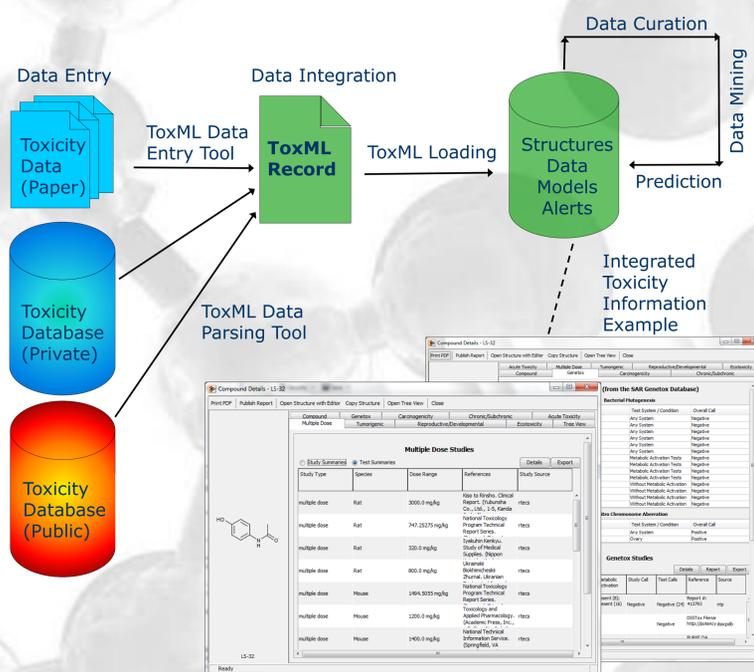


Abstract

With rapid scientific advances in toxicology and changing regulatory requirements, the number of toxicity datasets available for those wishing to share and communicate knowledge, or to use for data mining and modelling is continually expanding. The challenge with this ever-growing amount of data is that it exists in a multitude of different formats, depending on its source and original purpose. Different databases covering the same endpoint may omit or use widely differing conventions to structure and represent the same information. The issues of comparing or combining disparate data apply both to public and proprietary sources, and both historical and newly generated data. It is often laborious for individuals and groups to restructure their datasets in order to supply them to different investigators. The ToxML project addresses the need for a common data exchange standard that allows the representation and communication of this data in a well-structured electronic format to the user community.

ToxML is an open standard based on the Extensible Markup Language (XML) format. We describe the utility of ToxML as a common data exchange standard for toxicology information, the mechanism for its dissemination and its community-based development in the future. An example of how ToxML is currently implemented and used in transactions involving large quantities of repeat-dose study data is described. The standard is open and maintained by a curation team overseen by the ToxML organisation. The standard is published on a web site together with tools to view, edit and download it. Contributions from the user community to the ongoing evolution of the standard are facilitated in an open forum via a wiki on the web site.

Toxicity Information Workflow using ToxML



Scheme 1. An example workflow of how the exchange format can be used in collating, storing and modelling data from multiple sources.

What is ToxML?

ToxML is a community based project to promote an open data exchange standard that allows the communication of toxicological and related data in a well-structured electronic format. It is written in Extensible Markup Language (XML) format [1], and provides a means for representing and communicating toxicology data, tests, studies and compound information between individuals and groups creating and/or using such information.

Why use ToxML?

The current exchange of any generated toxicological data is reliant on the conversion of multiple formats of data by both the providers and/or interested recipients. The need for a common exchange standard has previously been identified [2]. For electronic records, a domain-specific mark up language provides methods to tag and recognise data elements in an efficient manner [3]. A common XML format allows multiple sources and formats of data to be communicated using a single mechanism without any other prerequisites. Each party only requires a single set of tools to read and write between their own internal formats and a centrally defined ToxML [Fig 1.]. The specification for ToxML is freely available and a mechanism has been established to allow the collaborative development of the standard to support new study designs.



Figure 1. How a common exchange standard removes the need for multiple conversions between different database systems.

ToxML online

The ToxML project is online at <http://toxml.org/>. The specification for the use of ToxML can be accessed and utilised on-line. The web site is currently managed by Lhasa Limited. To aid the use of ToxML, parser generators will be freely available.

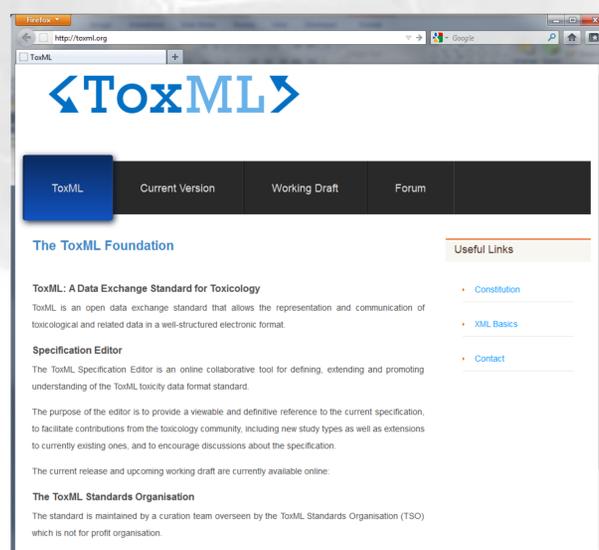


Figure 2. The current ToxML web site.

The ToXML Specification

The definition of how data should be structured in ToxML format is contained in the Specification. It consists of an XML Schema Definition (XSD) defining the toxicology schema and a controlled vocabulary that help ensure the consistency of usage when entering data in to the ToxML format.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<Compound>
  <Name>
    <Name type="chemName">R-(4-hydroxyphenyl)acetamide</Name>
  </Name>
  <ToxicityStudies>
    <BacterialMutagenesisStudies>
      <Study>
        <StudyCall>
          <Call>Muticrow</Call>
        </StudyCall>
      </Study>
    </BacterialMutagenesisStudies>
  </ToxicityStudies>
</Compound>
</Compound>
```

Figure 2. An example of a terse, but complete ToxML document.

Specification editor

The ToxML specification editor is the online collaborative tool for defining, extending and promoting understanding of the ToxML format. The editor provides views of the latest version of the published specification and the current working draft version. Registered users can edit the draft working version of the specification to modify the existing types or add new ones.

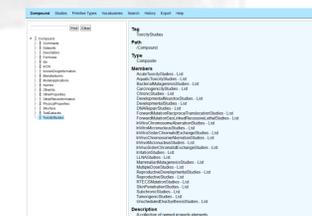


Figure 3. The ToxML Specification editor.

Discussion forum

This is a community based project and to facilitate discussion and contributions from the toxicology community a web forum has been established [Fig 4]. This will allow open discussion regarding the development and evolution of the specification.



Figure 4. The ToxML discussion forum.

The ToxML Standards Organisation

The standard is maintained by a curation team overseen by the ToxML Standards Organisation (TSO) which is not for profit. The aim of the TSO is to promote the use and development of ToxML as an international standard for the electronic exchange of toxicological and related data.

References:

- <http://www.w3schools.com/xml/default.asp>
- Richard, Ann M., and Williams, ClarLynda R. "Distributed structure-searchable toxicity (DSSTox) public database network: a proposal." Mutation Research. (2002) 499:27-52.
- Burgoon, Lyle D. "Clearing the standards landscape: the semantics of terminology and their impact on toxicogenomics." Toxicological Sciences. (2007) 99:403-412.

