Lhasa Limited
is a developer of expert knowledge based prediction software and chemical databases.

ToxML, a Data Exchange Standard with Content Controlled Vocabulary Used to Build Better (Q)SAR Models

Dr. MA Ali
Senior Scientist, Lhasa Limited

Tel: +44 (0)113 394 6020
Email: info@lhasalimited.org
Web: www.lhasalimited.org

© 2012 Lhasa Limited Registered Office: 22-23 Blenheim Terrace, Woodhouse Lane, Leeds, LS2 9HD, UK Registered Charity (290866)

Contents

- Background to the ToxML Project
  - Need for a data exchange standard
  - Fundamentals of ToxML
  - Existing exchange formats
- Structure of ToxML
  - Type of file
  - Toxicological data
  - Types of studies covered
- Wiki Website
  - Specification Editor
  - Contributing/Participating
  - Future Plans

Background

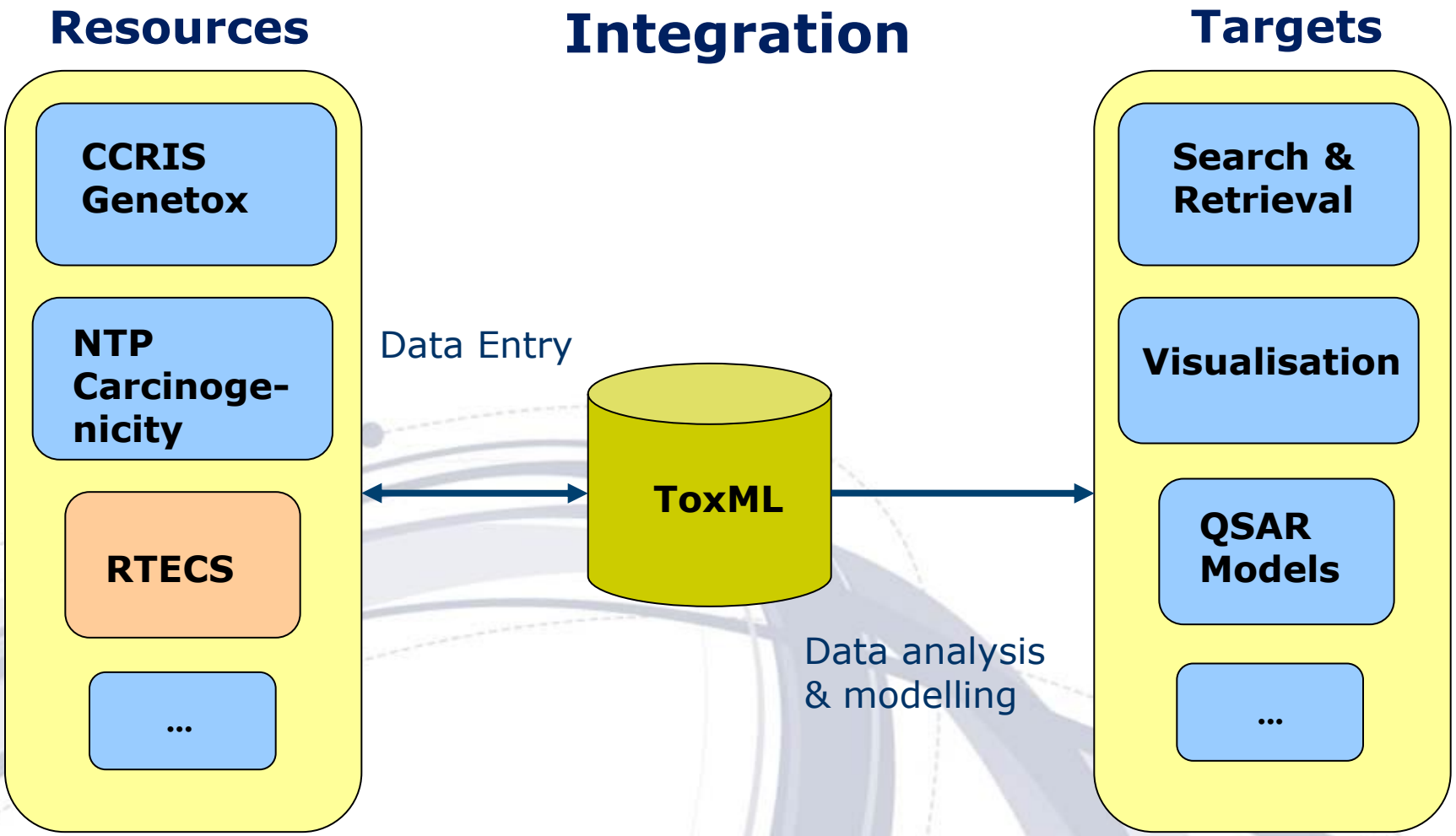
- The need for a data exchange standard

disparate

unified

- Submitting data to regulators
- Organisations communicate or share data
- Data mining, retrieval and visualization
- *in silico* predictive toxicology

Background



ToxML

Fundamentals of ToxML

- Original project created a preliminary exchange format

- Initiated by Leadscope & Lhasa
- Mainly developed by Leadscope in collaboration with US FDA
- Linked to the development of a data entry tool – freely available
- Covers single/repeat-dose toxicity studies & carcinogenicity studies
- XML – Extensible Markup Language:

  - hierarchical data structure

  - chemical compound serves as root node

  - supports both binary & non-binary data (numbers, text, images)
- Tox studies submitted as self contained data files
- Controlled/normalised vocabulary – ensures consistency

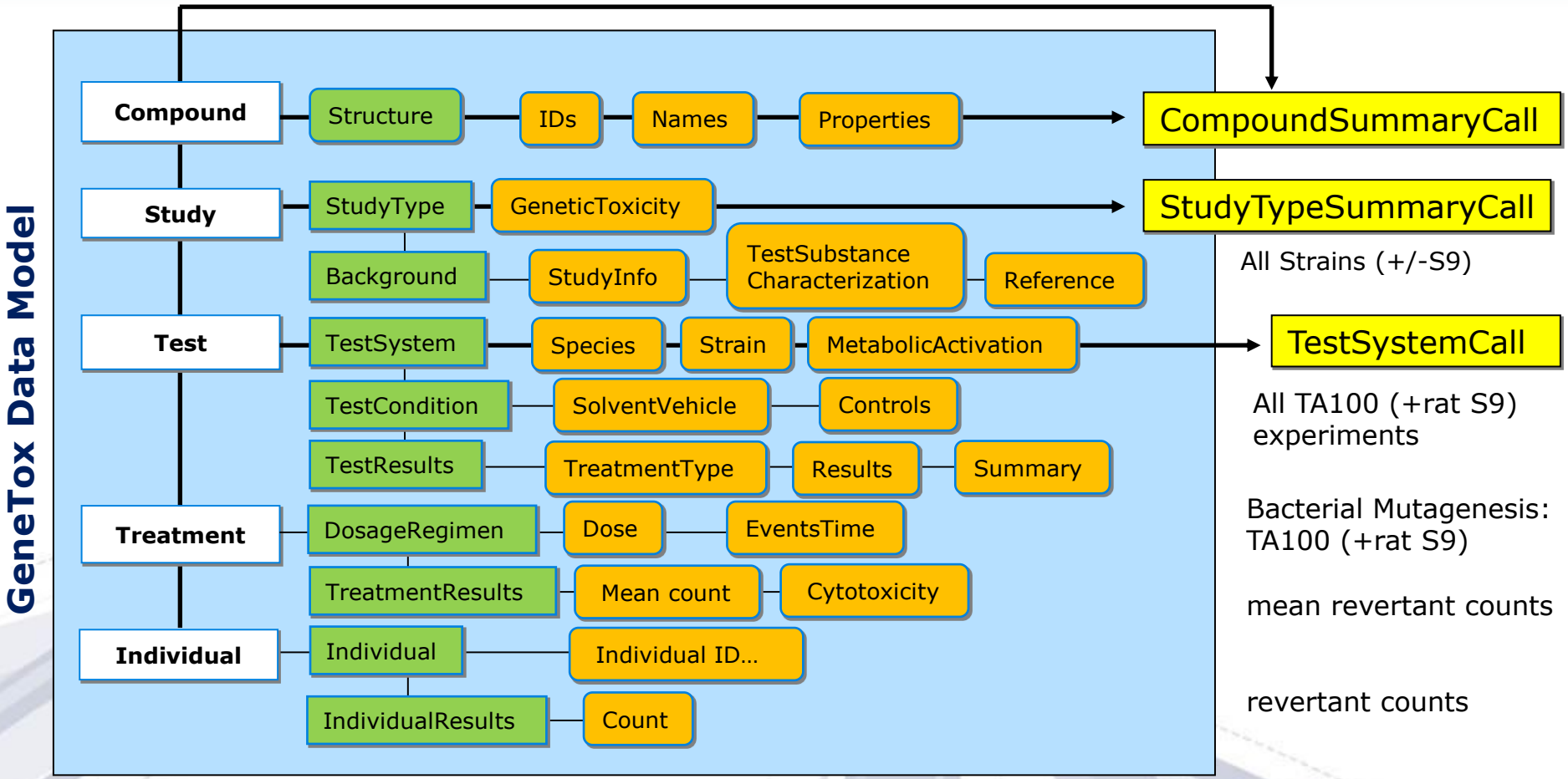
- Standard to be fully in the public domain

- Maintained by ToxML Standard Organisation (TSO)

Existing Exchange Formats

- SEND, REACH-IT
  - Designed primarily for submission of regulatory information
  - Developed for nonclinical data
- DSSTox
  - Mainly used for exchange of data via the EPA web site
  - Based on SD file format
- SDFiles
  - Chemical structure – well defined
  - All other data non-standardised
- CSV
  - Default format for flat files

Structure - overview

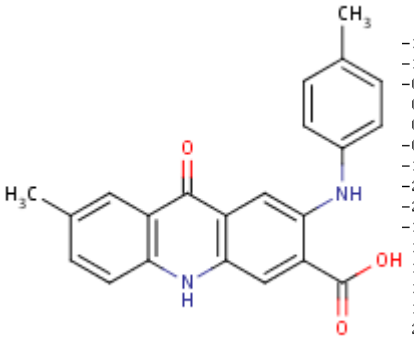


Toxicology Mechanisms and Methods 18:103–18, 2008

ToxML Record

Compound

- **Study**
  - Study type
  - Background
- **Test**
  - Test system
  - Test conditions
  - Test results
  - Treatments

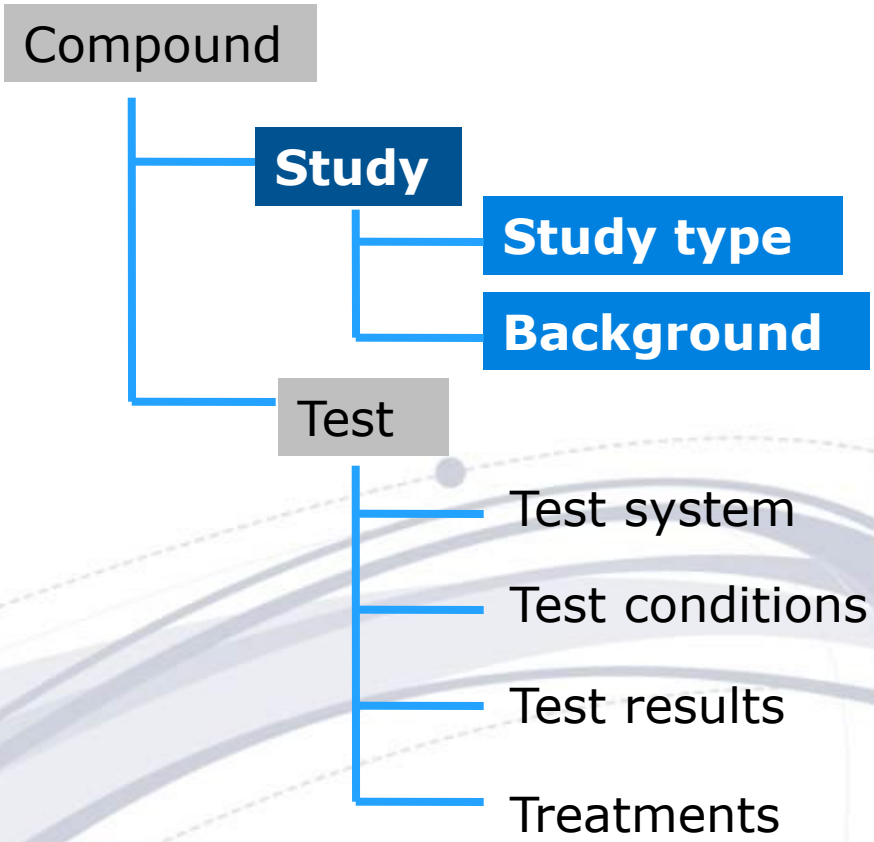



```
<Structure>
  <Molfile sources="stn">
    <ConnectionTable><![CDATA[ 27 30
-1.0717   -0.8250    0.0000 C   0  0  0
-1.0717   -1.6500    0.0000 C   0  0  0
-0.3572   -2.0625    0.0000 N   0  0  0
 0.3572   -1.6500    0.0000 C   0  0  0
 0.3572   -0.8250    0.0000 C   0  0  0
-0.3572   -0.4125    0.0000 C   0  0  0
-1.7862   -0.4125    0.0000 C   0  0  0
-2.5006   -0.8250    0.0000 C   0  0  0
-2.5006   -1.6500    0.0000 C   0  0  0
-1.7862   -2.0625    0.0000 C   0  0  0
 1.0717   -2.0625    0.0000 C   0  0  0
 1.7862   -1.6500    0.0000 C   0  0  0
 1.7862   -0.8250    0.0000 C   0  0  0
 1.0717   -0.4125    0.0000 C   0  0  0
 2.5006   -0.4125    0.0000 N   0  0  0
-0.3572    0.4125    0.0000 O   0  0  0
```

Field	Data
LS-181769	Leadscope Databases Created: 06-03-2011 00:00:00 Modified: 06-03-2011 00:00:00
Ids	
reg	LS-181769
cas	101904-52-3
fda	101904-52-3
Names	
chemName	2-Methyl-7-toluidinoquinacridone-6-carboxylic acid
synonym	TMAS; 3-Acridinecarboxylic acid, 9,10-dihydro-7-methyl-2-[(4-methylphenyl)amino]-9-oxo
chemName	3-Acridinecarboxylic acid, 9,10-dihydro-7-methyl-2-[(4-methylphenyl)amino]-9-oxo-(9CI) (CA INDEX NAME)
InChI	
InChI	1/C22H18N2O3/c1-12-3-6-14(7-4-12)23-20-10-16-19(11-17(20)22(26)27)24-18-8-5-13(2)9-15(18)21(16)25/h3-11,23H,1-2H3,(H,24,25)(H,26,27)
Toxicity Studies	
Datasets	
Text Datasets	

ToxML

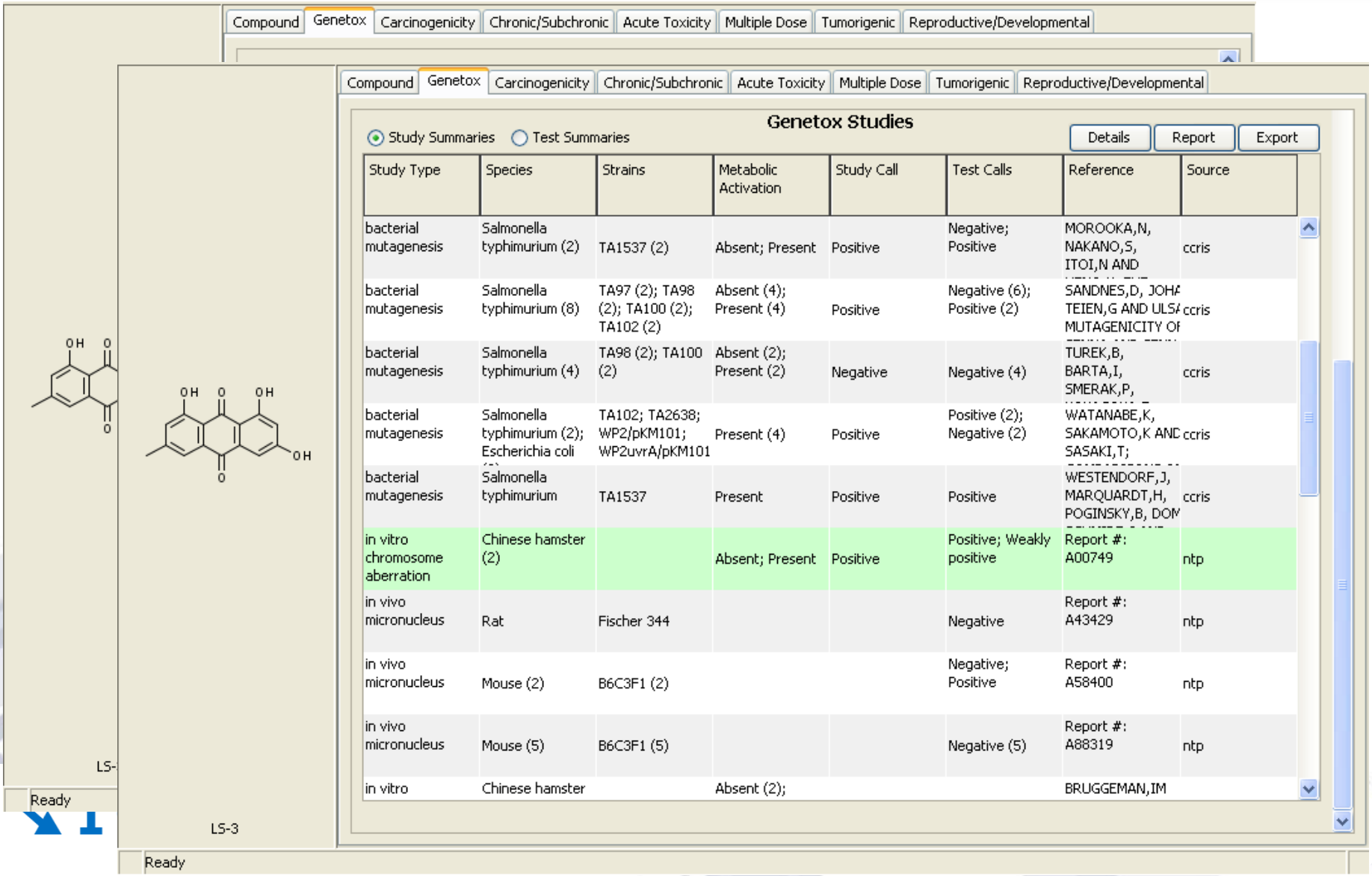
ToxML Record



Toxicity Studies	
Bacterial Mutagenesis Studies	
bacterial mutagenesis	cfsan-ofas
Study Type	bacterial mutagenesis
Study Calls	Negative
Background	
Regulatory Test Type	Regulatory original
Confidentiality	No
Physical Characteristics	
Test Matrix	Neat
Test Form	Parent
Percent Purity	technical quality
Percent Active	not reported
Physical State	Powder
Color	red
Color Index	None
Storage Condition	dark at 22 Celsius

ToxML

Integrated ToxML Record



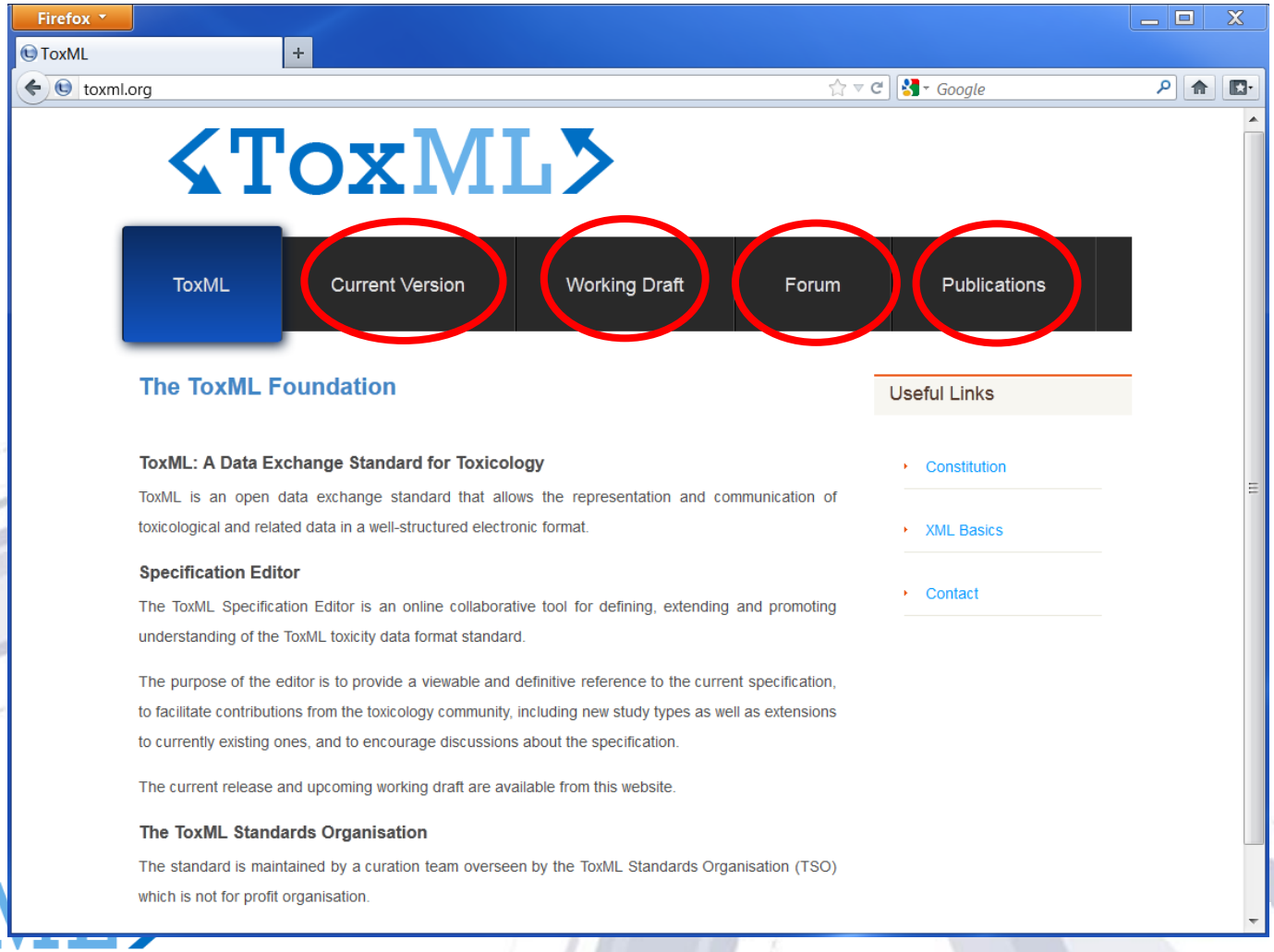
ToxML Data Types

- 27 Study/Data Types Modeled to Date

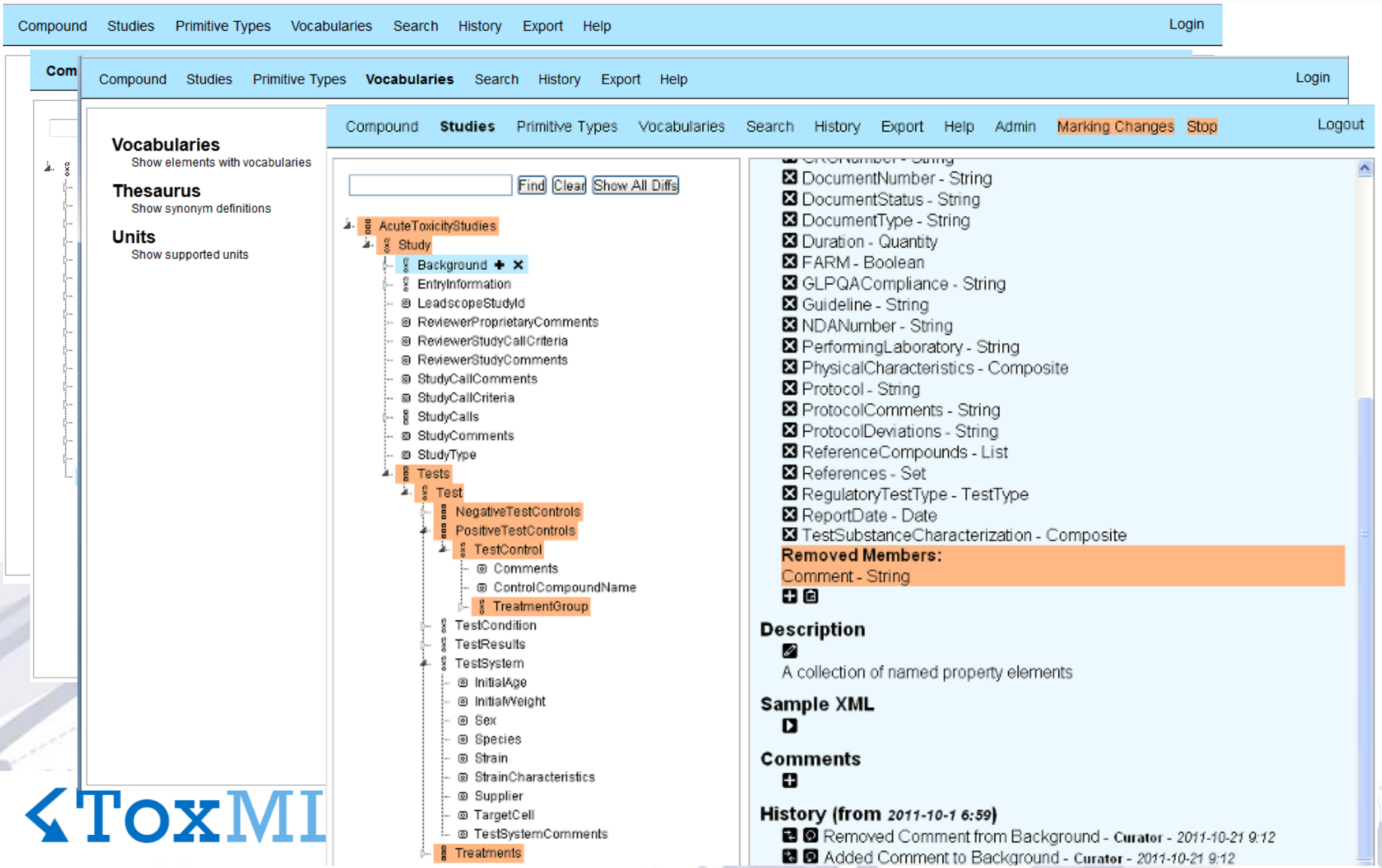
- **Genetic Toxicity**
  - Bacterial Mutagenesis
  - *in vitro* Chromosome Aberration
  - *in vitro* Micronucleus
  - *in vivo* Chromosome Aberration
  - *in vivo* Micronucleus
  - Mammalian Mutagenesis
- **Skin Penetration**
  - Bacterial Mutagenesis
- **Skin Sensitisation**
  - Local Nymph Node
- ***In Vivo* repeat dose studies**
  - Carcinogenicity
  - Chronic
  - Subchronic
  - Reproductive
  - Developmental
  - Developmental neurotoxicity
- ***In Vivo* Single Dose Studies**
  - Acute
- **Ecotoxicity**
  - Aquatic

Wiki Website

- http://www.toxml.org/



Specification Viewer and Editor



ToxMI

Future Use & Development of ToxML

- Extending ToxML

  Promote and encourage contributions through the ToxML Wiki site

- Efforts to link SEND and ToxML

  - Missing data fields, add animal data to test level information

  - Controlled vocabulary harmonization

  - Consolidate SEND study meta-data (SDTM) with ToxML

- Software tools to import/export from databases

  - Leadscope: parser generator & data entry tool kit

  - Lhasa: Vitic

  - eTOX

- New study types

  - Additional endpoints (contributors)

Summary

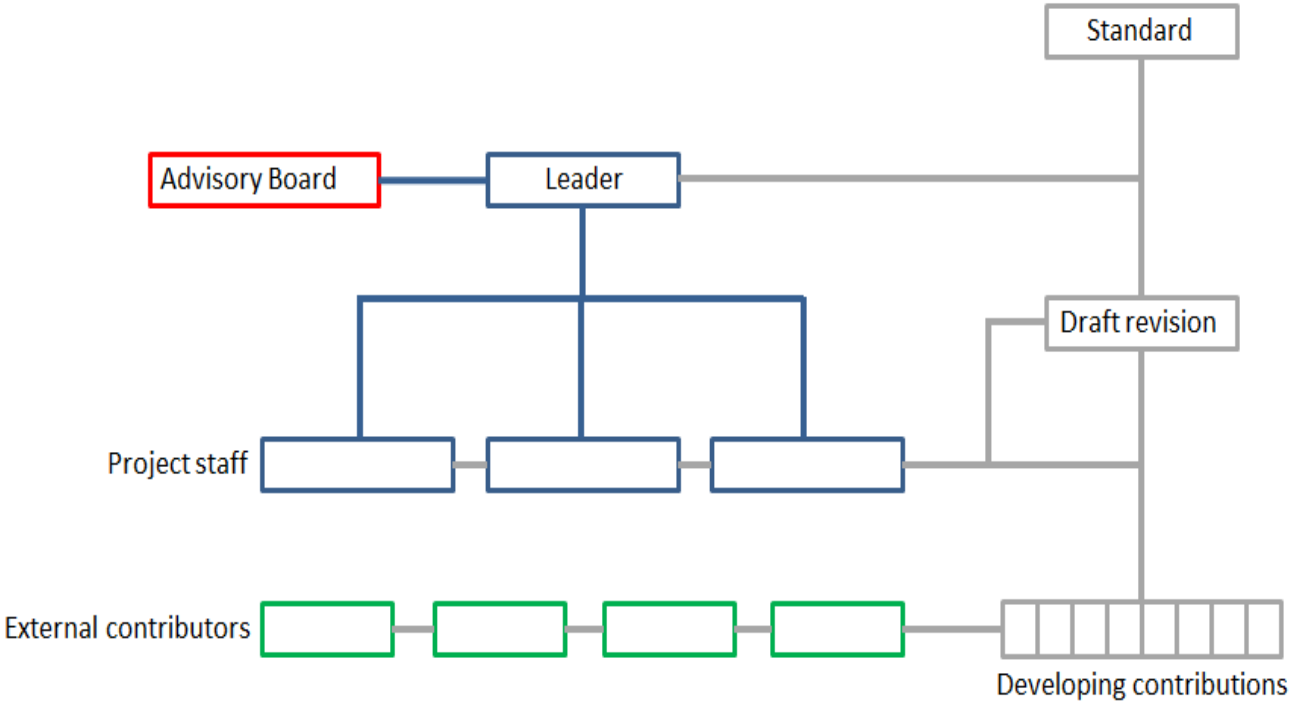
- Freely available data exchange standard
  - tools available now
  - designed to be extensible
  - 'community' approach: avoid chaotic expansion
  - any application capable of handling XML files
- Objective to make data capture & sharing easier
- Standard maintained by TSO, overseen by an advisory board
- Further information http://www.toxml.org

Acknowledgements

- Leadscope, Inc
  - Kevin Cross
  - David Bower
- Lhasa Limited
  - Mukesh Patel
  - David Wilkinson
- Chihae Yang (Altimira)
- Advisory Board:
  - Daniel Benz (US, FDA)
  - David Bower (Leadscope)
  - Philip Judson (Judson Consulting)
  - Nina Jeliazkova (IdeaConsult Ltd)
  - Igor Tetko (VCC Lab)
  - Mary Manibusan (EPA)
  - Bertrand Dagallier (OECD)
  - Shree Nath (PointCross)

Questions?

‹ToxML›



lhasa limited

Leadscope®