

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

◀ToxML▶

Leadscope[®]

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

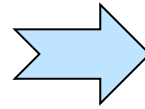
- Specification Editor
- Contributing/Participating
- Future Plans

Background

- ❖ The need for a data exchange standard



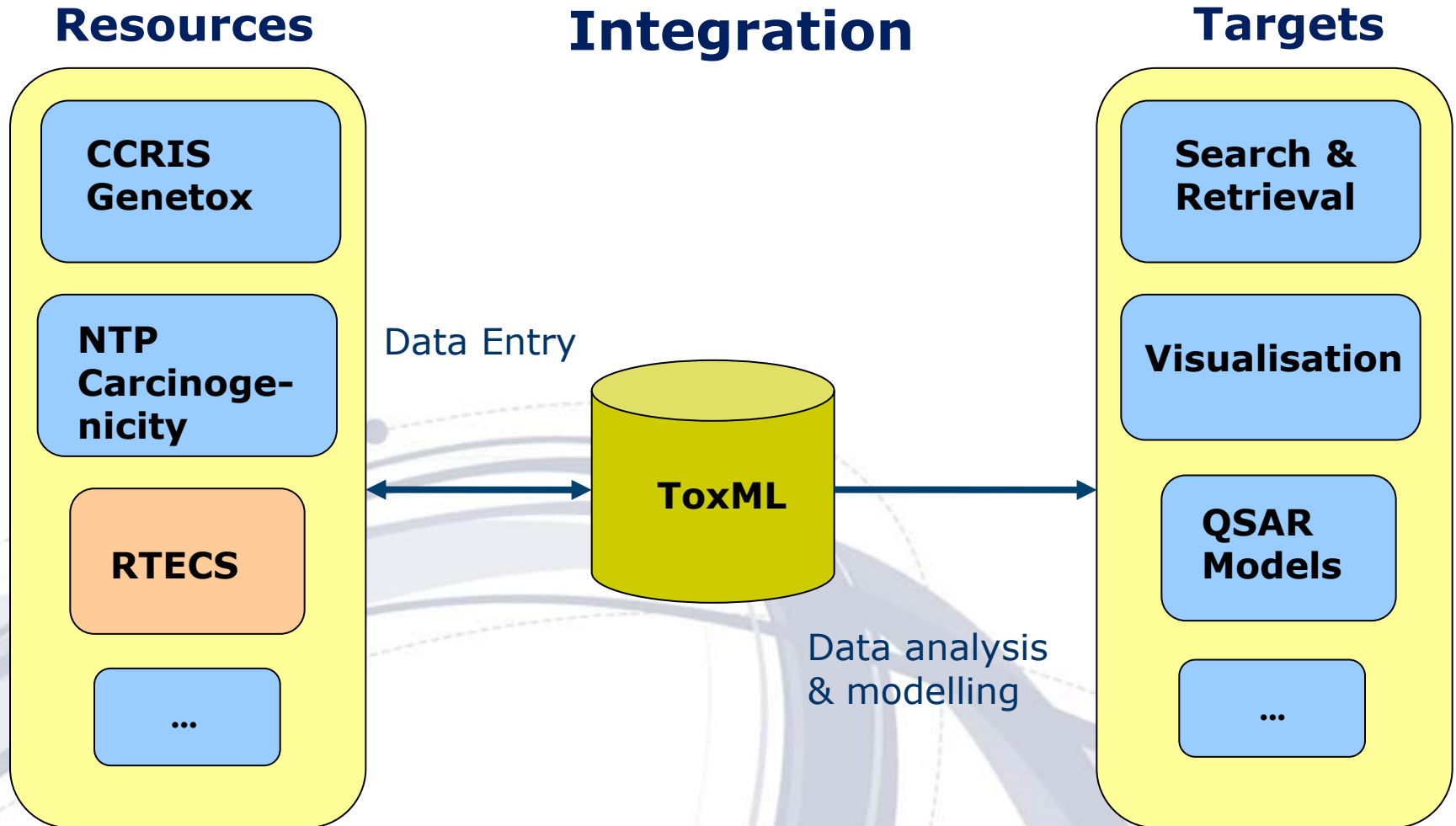
disparate



unified

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

Background



Fundamentals of ToxML

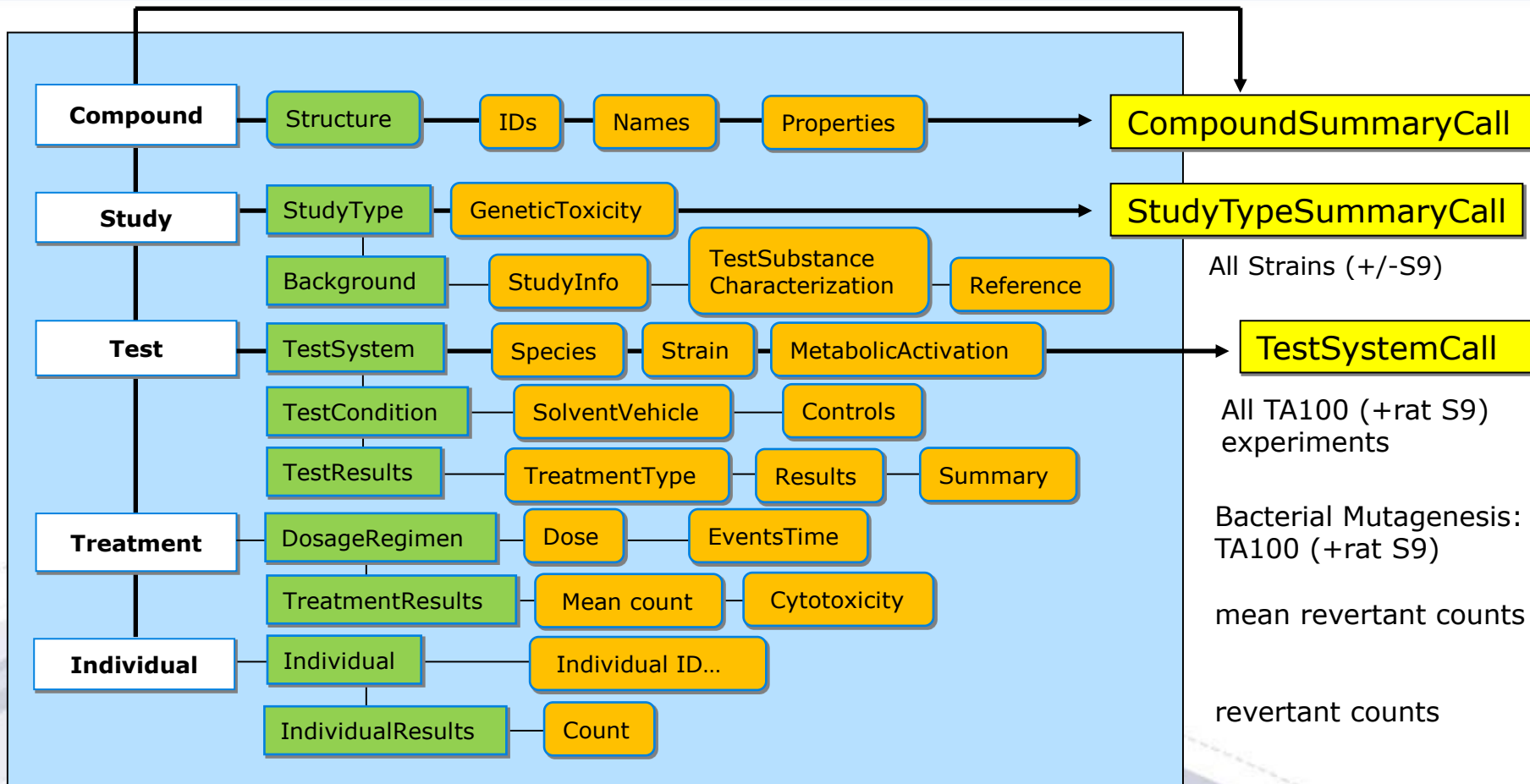
- ❖ Original project created a preliminary exchange format
 - Initiated by Leadscope & Lhasa
 - Mainly developed by Leadscope in collaboration with US FDA
 - Linked to the development of a data entry tool – freely available
 - Covers single/repeat-dose toxicity studies & carcinogenicity studies
 - XML – 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

GeneTox Data Model



Toxicology Mechanisms and Methods 18:103-18, 2008

ToxML Record

Compound

Study

Study type

Background

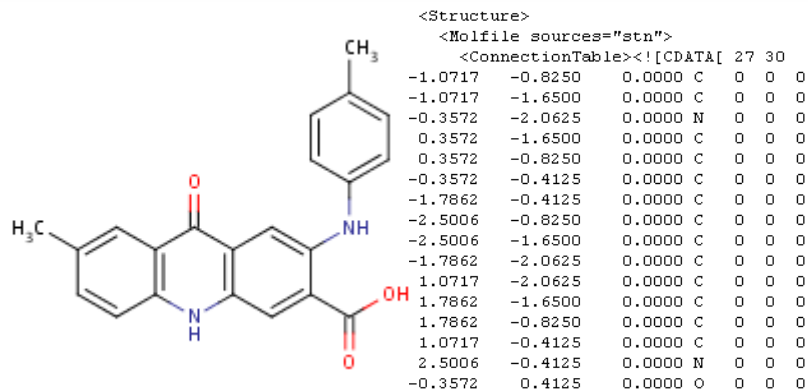
Test

Test system

Test conditions

Test results

Treatments



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 Record

Compound

Study

Study type

Background

Test

Test system

Test conditions

Test results

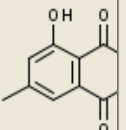
Treatments

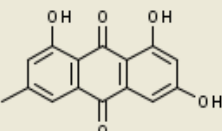
[-] Toxicity Studies	
[-] Bacterial Mutagenesis Studies	
[-] bacterial mutagenesis	cfsan-ofas
Study Type	bacterial mutagenesis
Study Calls	Negative
[-] Background	
Regulatory Test Type	Regulatory original
Confidentiality	No
[-] Physical Characteristics	
Test Matrix	Neat
Test Form	Parent
Percent Purity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

Integrated ToxML Record

Compound Genetox Carcinogenicity Chronic/Subchronic Acute Toxicity Multiple Dose Tumorigenic Reproductive/Developmental

Compound Genetox Carcinogenicity Chronic/Subchronic Acute Toxicity Multiple Dose Tumorigenic Reproductive/Developmental





Study Summaries
 Test Summaries

Details Report Export

Study Type	Species	Strains	Metabolic Activation	Study Call	Test Calls	Reference	Source
bacterial mutagenesis	Salmonella typhimurium (2)	TA1537 (2)	Absent; Present	Positive	Negative; Positive	MOROOKA,N, NAKANO,S, ITOI,N AND	ccris
bacterial mutagenesis	Salmonella typhimurium (8)	TA97 (2); TA98 (2); TA100 (2); TA102 (2)	Absent (4); Present (4)	Positive	Negative (6); Positive (2)	SANDNES,D, JOHA TEIEN,G AND ULSA MUTAGENICITY OF	ccris
bacterial mutagenesis	Salmonella typhimurium (4)	TA98 (2); TA100 (2)	Absent (2); Present (2)	Negative	Negative (4)	TUREK,B, BARTA,I, SMERAK,P,	ccris
bacterial mutagenesis	Salmonella typhimurium (2); Escherichia coli	TA102; TA2638; WP2/pKM101; WP2uvrA/pKM101	Present (4)	Positive	Positive (2); Negative (2)	WATANABE,K, SAKAMOTO,K AND SASAKI,T;	ccris
bacterial mutagenesis	Salmonella typhimurium	TA1537	Present	Positive	Positive	WESTENDORF,J, MARQUARDT,H, POGINSKY,B, DOY	ccris
in vitro chromosome aberration	Chinese hamster (2)		Absent; Present	Positive	Positive; Weakly positive	Report #: A00749	ntp
in vivo micronucleus	Rat	Fischer 344			Negative	Report #: A43429	ntp
in vivo micronucleus	Mouse (2)	B6C3F1 (2)			Negative; Positive	Report #: A58400	ntp
in vivo micronucleus	Mouse (5)	B6C3F1 (5)			Negative (5)	Report #: A88319	ntp
in vitro	Chinese hamster		Absent (2);			BRUGGEMAN,IM	

Ready
LS-3

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/>

Firefox

ToxML

toxml.org

Google

<ToxML>

ToxML

Current Version

Working Draft

Forum

Publications

The ToxML Foundation

ToxML: A Data Exchange Standard for Toxicology

ToxML is an open data exchange standard that allows the representation and communication of toxicological and related data in a well-structured electronic format.

Specification Editor

The ToxML Specification Editor is an online collaborative tool for defining, extending and promoting understanding of the ToxML toxicity data format standard.

The purpose of the editor is to provide a viewable and definitive reference to the current specification, to facilitate contributions from the toxicology community, including new study types as well as extensions to currently existing ones, and to encourage discussions about the specification.

The current release and upcoming working draft are available from this website.

The ToxML Standards Organisation

The standard is maintained by a curation team overseen by the ToxML Standards Organisation (TSO) which is not for profit organisation.

Useful Links

- ▶ [Constitution](#)
- ▶ [XML Basics](#)
- ▶ [Contact](#)

<ToxML>

Specification Viewer and Editor

Compound Studies Primitive Types Vocabularies Search History Export Help

Login

Com

Compound Studies Primitive Types **Vocabularies** Search History Export Help

Login

Compound **Studies** Primitive Types Vocabularies Search History Export Help Admin **Marking Changes** **Stop** Logout

Vocabularies

Show elements with vocabularies

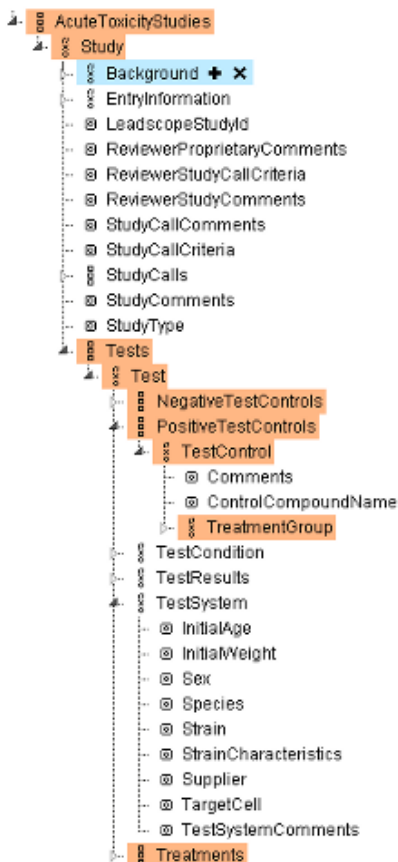
Thesaurus

Show synonym definitions

Units

Show supported units

Find Clear Show All Diffs



- DocumentNumber - String
- DocumentStatus - String
- DocumentType - String
- Duration - Quantity
- FARM - Boolean
- GLPQACompliance - String
- Guideline - String
- NDANumber - String
- PerformingLaboratory - String
- PhysicalCharacteristics - Composite
- Protocol - String
- ProtocolComments - String
- ProtocolDeviations - String
- ReferenceCompounds - List
- References - Set
- RegulatoryTestType - TestType
- ReportDate - Date
- TestSubstanceCharacterization - Composite

Removed Members:

Comment - String

Description

A collection of named property elements

Sample XML

Comments

History (from 2011-10-1 6:59)

- Removed Comment from Background - Curator - 2011-10-21 9:12
- Added Comment to Background - Curator - 2011-10-21 9:12

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 Jeliaskova (IdeaConsult Ltd)
 - Igor Tetko (VCC Lab)
 - Mary Manibusan (EPA)
 - Bertrand Dagallier (OECD)
 - Shree Nath (PointCross)

Questions?

↳ ToxML ↲

