Data Collation and Curation for the COSMOS Database

Longzhi Yang¹, Yang Lan¹, Kirk Arvidson², Mark TD Cronin³, Steven J Enoch³, Johann Gasteiger⁴, Dimitar Hristozov², Judith C Madden⁵, Daniel Neagu¹, Andrea-N Richarz³, James F Rathman⁵, Mick J Ridley¹, Oliver Sacher⁴, Christof H Schwab⁴, Aleksey Tarkhov⁴, Chihae Yang⁵

¹ University of Bradford, England, ² US FDA CFSAN, College Park, MD, USA, ³ Liverpool John Moores University, England, ⁴ Molecular Networks GmbH, Erlangen, Germany, ⁵ Altamira LLC, Columbus, OH, USA

L.Yang8@Bradford.ac.uk

Introduction and Aims
The management and sharing of chemical, biological and toxicity data play a central role within the COSMOS project. The data are collected, stored and managed to support other work packages in their predictive modelling tasks. The harvested data will be stored in a freely available toxicity database which is organised in a relational database model in an optimised way. The resulting databases will be properly quality-controlled, curated, governed, and secured.

Results
- Identified available toxicity data sources;
- Developed a data schema;
- Designed the data entry system;
- Deployed “proof of concept” database;
- Analysed a data authorisation model;
- Proposed a data quality-control, curation and data governance strategy.

Dimensions for toxicological data quality

Decision domains of data governance for toxicological database

Conclusions
The requirements for the COSMOS database have been specified, a data quality-control, curation and governance strategy has been proposed and the system has been designed. The next steps will be the implementation into products for data collection to serve the COSMOS project within the SEURAT-1 cluster.

Acknowledgements
The research leading to these results has received funding from the European Community’s Seventh Framework Program (FP7/2007-2013) under grant agreement n° 266835 and from Colipa.

Key References