Facilitating Data Curation: a Solution Developed in the Toxicology Domain

Christophe Debruyne[1,3], Jonathan Riggio[1], Emma Gustafson[2],
Declan O'Sullivan[3], Mathieu Vinken[2], Tamara Vanhaecke[2], and Olga De Troyer[1]

[1]WISE Lab, Vrije Universiteit Brussel, Belgium
[2]Research Group of In Vitro Toxicology and Dermato-Cosmetology (IVTD), Vrije Universiteit Brussel, Belgium
[3]ADAPT Centre, Trinity College Dublin, Ireland

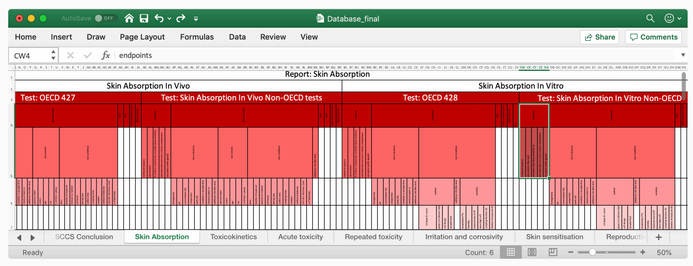
Context

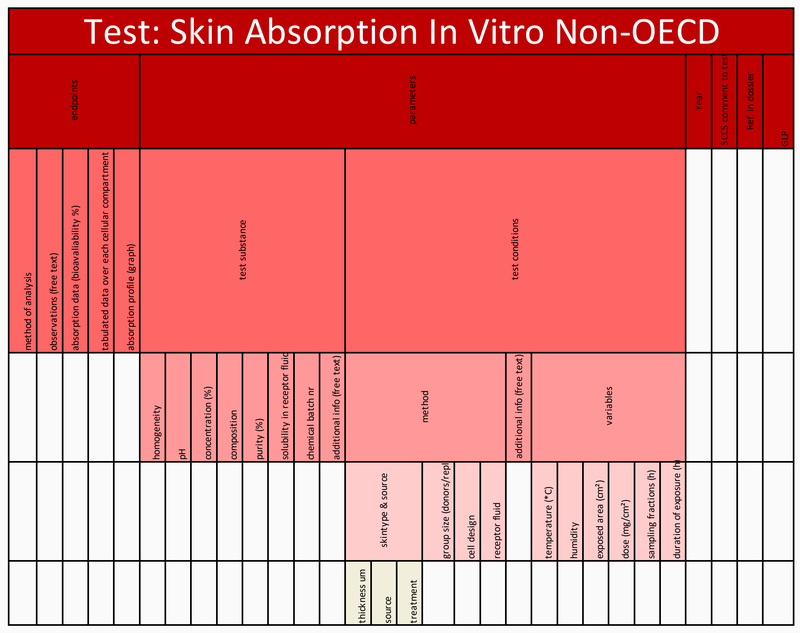
- The goal of toxicology is to understand the adverse effects of chemical compounds or physical agents on living organisms.
- The IVTD Group in the VUB aims to collect safety testing data of cosmetic ingredients available in publicly available safety evaluation reports with the goal of creating a knowledge base.

Problem

- Subject matter experts currently rely on spreadsheets, which causes problems for: data curation, data analysis, and data exploration.
- Semantic Technologies can overcome these challenges and furthermore allows data to be linked with external datasets.
- But semantic technologies are not easily accessible to these subject matter experts [1]
- How can we facilitate the creation of a knowledge base for the available safety evaluation reports by subject matter experts in the domain of toxicology?

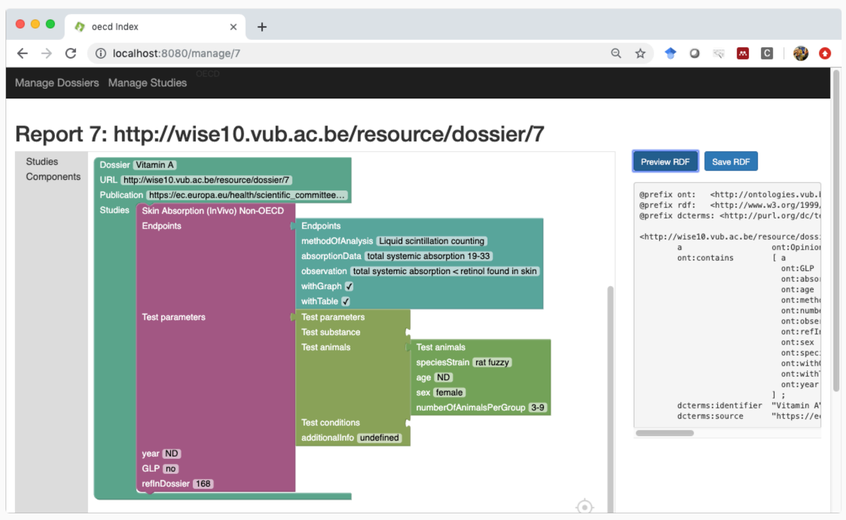
Structure of a Spreadsheet





Problem

- An interface metaphor [2] is drawing upon the knowledge of familiar concepts to facilitate learning and using a system.
- The Jigsaw Metaphor is proven successful for other tasks.
- In our proposed solution, the Jigsaw puzzle pieces guide subject matter experts in creating valid data entries.
- The prototype is built on top of Google Blockly for the metaphor and Apache Jena for the knowledge base.

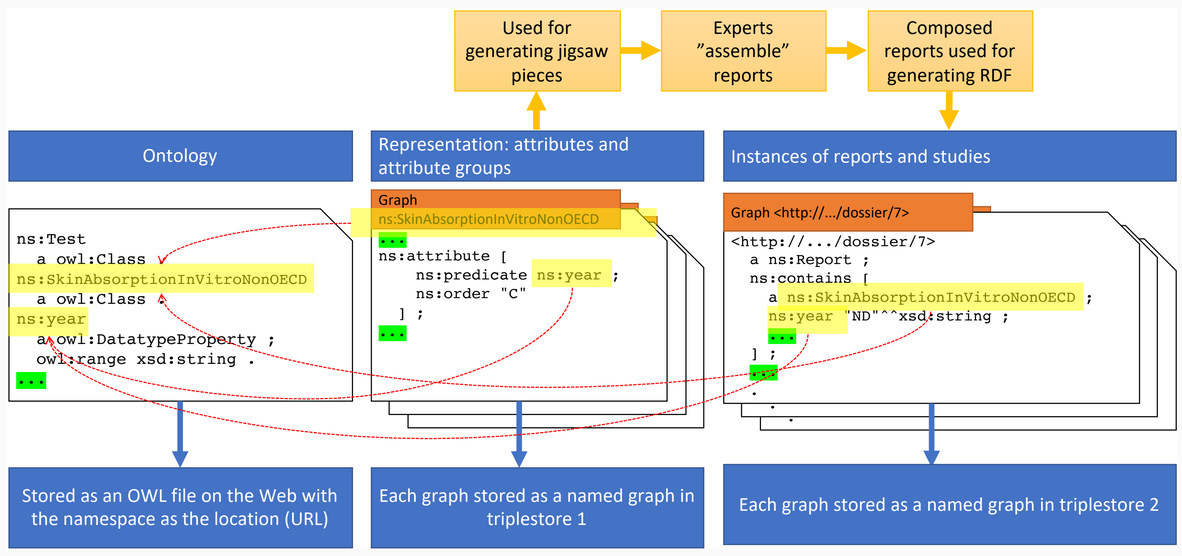


Knowledge Organization

The knowledge architecture consists of:

- A lightweight ontology containing few axioms for extensibility.
- The structure of jigsaw pieces in separate named graphs.
- The knowledge base containing the data of safety evaluation reports–one graph per report.

Both the creation of the ontology and the structure of the jigsaw pieces are currently driven by the existing spreadsheets.



Conclusions and Future Work

We proposed the adoption of a jigsaw metaphor to facilitate data curation by subject matter experts. Future work is twofold:

- Use of the tool by subject matter experts for foreseen in 2020
- Adoption of the jigsaw metaphor defining and maintaining the various blocks

References

1. L. McKenna, C. Debruyne, and D. O'Sullivan, "Understanding the position of information professionals with regards to linked data: A survey of libraries, archives and museums," in Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, JCDL 2018, Fort Worth, TX, USA, June 03-07, 2018, J. Chen et al., Eds. ACM, 2018, pp. 7–16.
2. T. D. Erickson, "Working with interface metaphors," in Readings in Human-Computer Interaction. Elsevier, 1995, pp. 147–151.

Acknowledgements

This work was partially supported by the funds of Cosmetics Europe. Christophe Debruyne is supported by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

WEB & INFORMATION SYSTEMS ENGINEERING
IN VITRO TOXICOLOGY & DERMATO-COSMETOLOGY RESEARCH GROUP
VRIJE UNIVERSITEIT BRUSSEL
Cosmetics Europe the personal care association
Engaging Content Engaging People
Trinity College Dublin Coláiste na Tríonóide, Baile Átha Cliath The University of Dublin