Faculty of Health, Medicine and Life Sciences

Enhancing the AOP-Wiki usability and accessibility with semantic web technologies

Marvin Martens[A], Trevor A Levey[B], Holly M Mortensen[B], Chris T Evelo[A,C], Egon L Willighagen[A]

[A] Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, The Netherlands
[B] US Environmental Protection Agency, National Health and Environmental Effects Laboratory, Research Cores Unit, Research Triangle Park, NC 27709, USA.
[C] Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, The Netherlands

Introduction

There is a need for faster and more efficient use of existing data to assemble effective assessment strategies for the ever growing number of chemicals. Therefore, a framework to organize existing mechanistic information, the Adverse Outcome Pathway (AOP), was introduced. [1] The main repository for such AOPs is the AOP-Wiki. However, it is challenging to automatically and systematically parse, filter, and use its captured knowledge.

Objective

We explored the use of semantic web technologies to link the AOP-Wiki with chemical and biological databases, allowing more detailed exploration of the database, thereby better supporting risk assessment workflows.

Methods

- The AOP-Wiki XML was parsed and converted into Resource Description Framework (RDF) using a custom Python script.
- All parts of the AOP-Wiki, and the relationships between them, were semantically annotated using ontologies and standard vocabularies such as AOPO, CHEMINF, RDFS, and DC.
- The AOP, KE, KER and Stressor were registered in the MIRIAM Registry [2] and the resulting resolvable Uniform Resource Identifiers (URIs), along with the ones for chemicals and text-mapped genes, were integrated in the RDF.
- The identifier mapping service BridgeDb [3] was used to add matching identifiers for chemicals and genes of twelve additional databases.
- Similarly, parts of the AOP-DB [4,5] were converted in RDF as part of the OpenRiskNet implementation challenge.

Results

This exercise resulted in a conversion of the complete AOP-Wiki into an RDF schema (Figure 1) that includes over 66,000 triples. 16 ontologies were used for the semantic annotations, and over 6,000 persistent identifiers of 20 types are included. Also, the AOPO was used to annotate over 4,000 components, and the addition of the AOP-DB RDF introduced over 5,000 diseases, 3,000 pathways, and 1,000 ToxCast assays related to relevant genes.

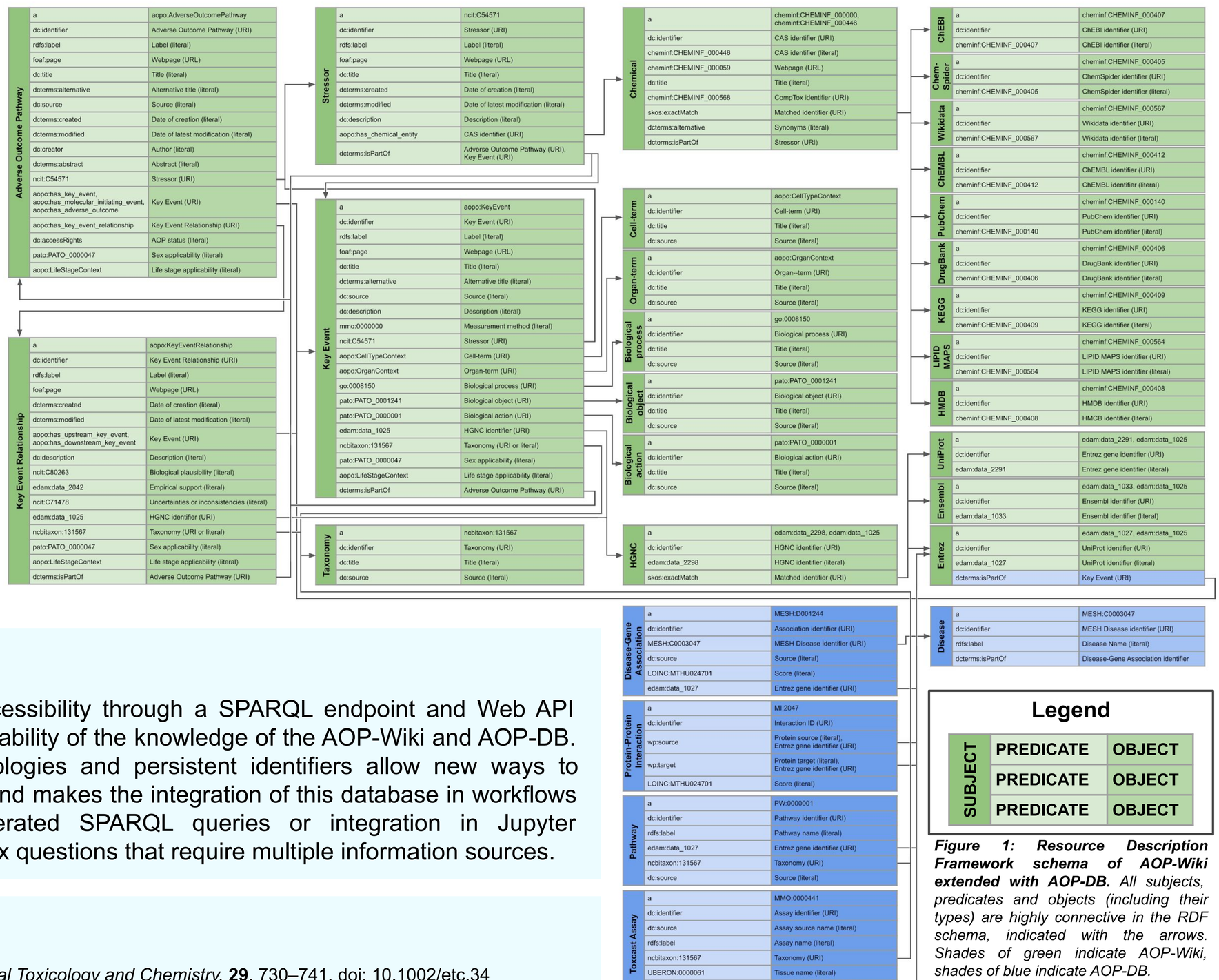

Figure 1: Resource Description Framework schema of AOP-Wiki extended with AOP-DB. *All subjects, predicates and objects (including their types) are highly connective in the RDF schema, indicated with the arrows. Shades of green indicate AOP-Wiki, shades of blue indicate AOP-DB.*

Conclusion

The created RDF and its accessibility through a SPARQL endpoint and Web API assist in the expansion and usability of the knowledge of the AOP-Wiki and AOP-DB. Furthermore, the use of ontologies and persistent identifiers allow new ways to explore the AOP knowledge, and makes the integration of this database in workflows possible. For example, federated SPARQL queries or integration in Jupyter notebooks can answer complex questions that require multiple information sources.

References

[1] Ankley, GT et al. (2010). *Environmental Toxicology and Chemistry.* **29**, 730–741. doi: 10.1002/etc.34
[2] Juty, N et al. (2012). *Nucleic Acids Research*, **40**, 580–586. doi: 10.1093/nar/gkr1097
[3] van Iersel, MP et al. (2010). *BMC Bioinformatics* **11**. doi: 10.1186/1471-2105-11-5
[4] Pittman, ME et al. (2018). *Toxicology and Applied Pharmacology* **343**, 71-83. doi: 10.1016/j.taap.2018.02.006
[5] Mortensen, HM et al. (2018). *Mammalian Genome* **29**, 190-204. doi: 10.1007/s00335-018-9738-7

Correspondence to:
Marvin Martens
ORCID: 0000-0003-2230-0840
marvin.martens@maastrichtuniversity.nl
@marvinm2

Maastricht University
Department of Bioinformatics – BiGCaT
http://www.bigcat.unimaas.nl
T +3143 388 1343

This project has received funding from the European Union's Horizon 2020 research and innovation programme project EU-ToxRisk under grant agreement No. 681002 and EINFRA-22-2016 programme project OpenRiskNet under grant agreement No. 731075.

The AOP-DB is supported by the EPA National Program in Chemical Safety and Sustainability and contributes to the AOP Discovery and Development; Informatics, Synthesis, and Integration; and Chemical Safety Analytics Project areas.