Maastricht University | Maastricht UMC+ | BiGCaT bioinformatics | NUTRIM

Faculty of Health, Medicine and Life Sciences

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

Marvin Martens[A], Chris T Evelo[A,B], Egon L Willighagen[A]

[A] Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, The Netherlands

[B] 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 and allowing **more detailed exploration of the database**, thereby better supporting risk assessment workflows.

Results

The complete AOP-Wiki was converted into an RDF schema (**Figure 1**) that includes over 98,000 unique triples, based around 280 AOPs (**Table 1**). Eight metadata vocabularies and seventeen domain-specific ontologies were used for the semantic annotations, and over 77,000 persistent identifiers of twenty types are included. Also, the AOPO was used to annotate over 10,000 components. The AOP-Wiki RDF can be used for rapid extraction of AOP-Wiki content through a SPARQL endpoint (**Figure 2**).

Element	# of subjects	# of properties
Adverse Outcome Pathway	280	8184
Key Event	1080	15556
Key Event Relationship	1259	15881
Stressor	454	4821
Chemical	293	9869
Taxonomy	88	373
Cell-term	59	261
Organ-term	61	268
Biological Process	353	1417
Biological Object	358	2046
Biological Action	0	0

Table 1: AOP-Wiki RDF statistics. *The number of subjects and properties for each of the eleven core elements of the AOP-Wiki.*

Conclusion

The created RDF and its accessibility through a SPARQL endpoint assist in the **expansion and usability** of the knowledge of the AOP-Wiki. 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.

Methods

The AOP-Wiki XML was parsed and converted into **Resource Description Framework (RDF)**. While doing that, all properties and objects (if applicable) were **semantically annotated** using ontologies and standard vocabularies such as the AOP Ontology, chemical information ontology, RDF Schema, and Dublin Core. Furthermore, resolvable **Internationalized Resource Identifiers (IRIs)** [2] were placed for AOP-specific subjects, chemicals, and proteins, and **additional molecular identifiers** were added from nine chemical databases and four gene/protein databases through identifier mapping with BridgeDb [3].

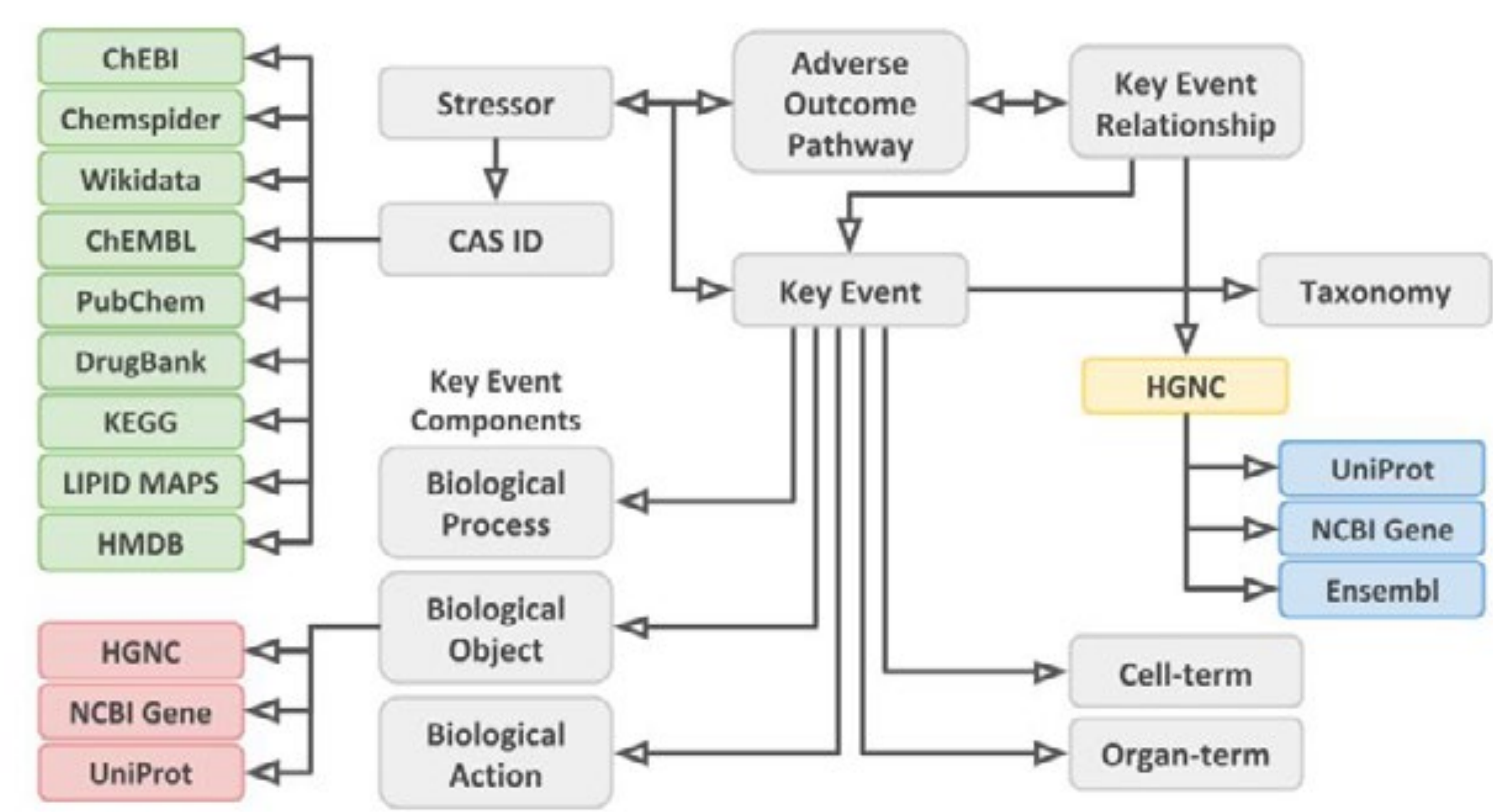

Figure 1: General overview of the AOP-Wiki RDF scheme. *Arrows show the directional relationships described in the RDF. Grey boxes are the basic elements of the AOP-Wiki. Green boxes indicate added chemical IDs using BridgeDb. Red boxes indicate added gene/protein IDs using Protein Ontology mapping. The yellow box indicates the text-mapped gene IDs and the blue boxes indicate the added gene/protein IDs mapped from the text-mapped gene IDs using BridgeDb.*

SPARQL Query:

```
SELECT DISTINCT ?AOname ?MIEname ?CASID ?ChemicalName  WHERE {
    ?AO a aopo:KeyEvent ; dc:title ?AOname .
    ?AOP a aopo:AdverseOutcomePathway ;  aopo:has_adverse_outcome ?AO ;
aopo:has_molecular_initiating_event ?MIE.
    ?MIE dc:title ?MIEname .
    ?stressor a nci:C54571 ; dcterms:isPartOf ?MIE .
    ?chemical a cheminf:CHEMINF_000000 ; dcterms:isPartOf ?stressor ; dc:title
?ChemicalName ; cheminf:CHEMINF_000446 ?CASID.
FILTER regex(str(?AOname), "Impairment, Learning and memory") }
```

AOname	MIEname	CASID	ChemicalName
"Impairment, Learning and memory"	"Binding of antagonist, NMDA receptors"	"7439-92-1"	"Lead"
"Impairment, Learning and memory"	"Inhibition, Na+/I- symporter (NIS)"	"14797-73-0"	"Perchlorate"
"Impairment, Learning and memory"	"Inhibition, Na+/I- symporter (NIS)"	"14797-55-8"	"Nitrate"
"Impairment, Learning and memory"	"Inhibition, Na+/I- symporter (NIS)"	"302-04-5"	"Thiocyanate"
"Impairment, Learning and memory"	"Binding, Thiol/seleno-proteins involved in protection against oxidative stress"	"79-06-1"	"Acrylamide"
"Impairment, Learning and memory"	"Binding, Thiol/seleno-proteins involved in protection against oxidative stress"	"115-09-3"	"Methylmercuric(II) chloride"
"Impairment, Learning and memory"	"Binding, Thiol/seleno-proteins involved in protection against oxidative stress"	"51312-24-4"	"Mercury chloride"
"Impairment, Learning and memory"	"Binding of agonist, Ionotropic glutamate receptors"	"14277-97-5"	"L-Domoic acid"

Figure 2: Example of a SPARQL query against the AOP-Wiki Resource Description Framework. *The SPARQL query extracts all chemicals that are linked to the Molecular Initiating Events of AOPs with an Adverse Outcome of interest, which can be queried by free text, highlighted in yellow. The output table indicates that there are eight chemicals in the AOP-Wiki that arelinked to learning and memory impairment.*

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

OpenRiskNet RISK ASSESSMENT E-INFRASTRUCTURE | EUTOXRISK

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

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.