





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.

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

Objective

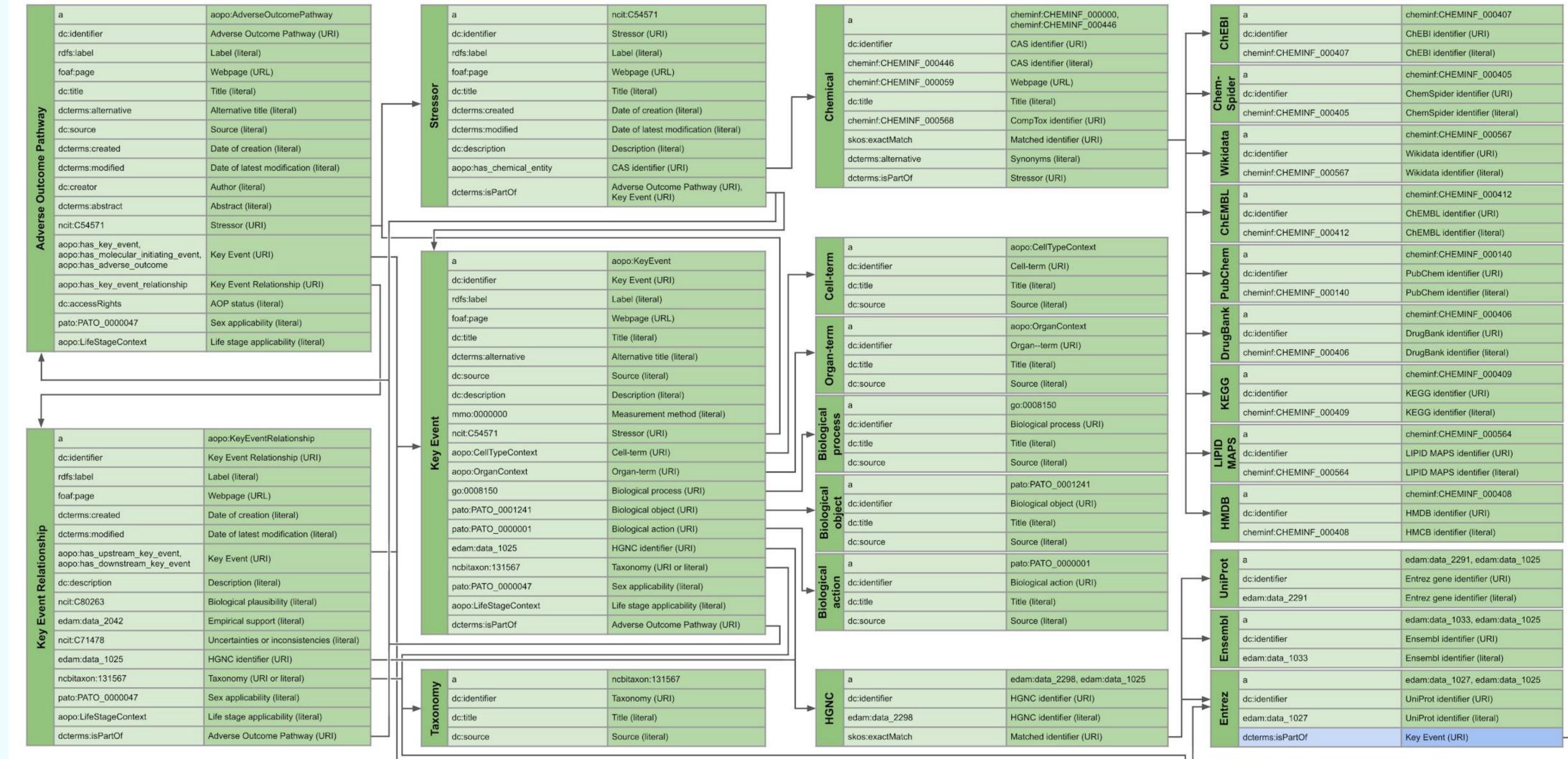
We explored the use of semantic web technologies to link the AOP-Wiki with chemical and biological databases, allowing more detailed exploration of the database, thereby better supporting risk assessment workflows.

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



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

A a cidentifier (UR) a detam.data_2298, edam.data_1025 de:dentifier UR) de:dentifier (UR) de:dentifier UR) de:dentifier (UR)						-						
0 dc.source The (theraf) 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.source edam.data_102 0 dc.source MESH Doutset 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.sourc	-	-	а	pato:PATO_0000001		ot	а		edam:da	ata_2291, edam:data_1025		
0 dc.source The (theraf) 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.source edam.data_102 0 dc.source MESH Doutset 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.source Source (theraf) 0 dc.sourc		gice	dc:identifier	Biological action (URI)		niP	dc:identifier		Entrez gene identifier (URI)			
dcisource (Reni) dcisource (Reni) dcisource (Reni) a MESH-D001244 a MESH-D001244 dcisource (Reni) a MESH-D000017 MESH-D000247 dcisource (Reni) a MESH-D001244 dcisource (Reni) a MESH-D00017 MESH-D000217 MESH-D000217 MESH-D000017 Accounce (Reni) a MC00001 a MC0001 Accounce (Reni) a MC00001	-	acti	dc:title	Title (literal)		D	edam:data_2291		Entrez gene identifier (literal)			
a edam.data_2288, edam.data_1025 dc:dentifier HGNC identifier (URi) edam.data_2288 HGNC identifier (URi) scs.swactMatch Matched identifier (URi) dc:dentifier UniProt identifier (URi) dc:dentifier Matched identifier (URi) dc:dentifier Assocution identifier (URi) dc:dentifier Assocution identifier (URi) dc:dentifier Assocution identifier (URi) dc:dentifier MESH Docuse identifier (URi) dc:dentifier Interaction ID (URI) word Entre gene identifier (URI) word ME207	0	٥	dc:source	Source (literal)		Idr	a		edam:data_1033, edam:data_1025			
a edam.data_2288, edam.data_1025 dc:dentifier HGNC identifier (URi) edam.data_2288 HGNC identifier (URi) scs.swactMatch Matched identifier (URi) dc:dentifier UniProt identifier (URi) dc:dentifier Matched identifier (URi) dc:dentifier Assocution identifier (URi) dc:dentifier Assocution identifier (URi) dc:dentifier Assocution identifier (URi) dc:dentifier MESH Docuse identifier (URi) dc:dentifier Interaction ID (URI) word Entre gene identifier (URI) word ME207					` →	Ensen	dc:identifier		Ensembl identifier (URI)			
Yey dc:dentifier HGNC:dentifier (UR1) dc:dentifier UniProt identifier (UR1) dc:dentifier UniProt identifier (UR1) dc:dentifier MESH:D001244 a MESH:D001244 dc:dentifier Association identifier (UR1) dc:dentifier Association identifier (UR1) dc:dentifier Association identifier (UR1) dc:dentifier Association identifier (UR1) dc:dentifier MESH:D001244 dc:dentifier MESH:D00124 dc:dentifier MESH:D00124 dc:dentifier MESH:D00124 dc:dentifier MESH:D00124 dc:dentifier MESH:D00124<									Ensembl identifier (literal) edam:data_1027, edam:data_1025			
stocexactMatch Matched identifier (UR)) dcdemtsilsPartOf key Event (UR)) a MESH-D001244 a MESH-D000047 dc:dontifier Association identifier (UR)) dc:dontifier MESH-D003047 MESH-D003047 MESH-D001244 MESH-D003047 dc:dontifier Association identifier (UR)) dc:dontifier MESH-D003047 MESH-D003047 MESH-D001244 MESH-D003047 MESH-D003047 MESH-D003047 MESH-D001244 Disease-Identifier (UR)) Disease-Gene Association identifier (UR) dc:dontifier Interaction ID (UR) Disease-Gene Association identifier (UR) Disease-Gene Association identifier (UR) wptarget a Mi2047 Disease-Gene Association identifier (UR) wptarget Protein source (iteral) Entrez gene identifier (UR) PREDICATE OBJECT vptarget a PProtein source (iteral) Protein source (iteral) Disease-Gene Association identifier (UR) edemtifier Pathway identifier (UR) Entrez gene identifier (UR) PREDICATE OBJECT vptarget a PProtein source (iteral) Figure 1: Resource Description Framework schema of AOP-V			а	edam:data_2298, edam:data_1025								
skosexactMatch Matched identifier (URI) a MESH-D001244 dc:dentifier Association identifier (URI) MESH-C0003047 MESH Doesse identifier (URI) dc:dentifier IMESH Doesse identifier (URI) dc:dentifier Interaction ID (URI) mporting at Mi2047 dc:dentifier Interaction ID (URI) mporting at Protein source (iteral) potarget Protein source (iteral) cidentifier Protein source (iteral) cidentifier Protein source (iteral) cidentifier Protein source (iteral) dc:dentifier Pathway identifier (URI) rd:ds:dstall Potone source (iteral) dc:dentifier Pathway name (iteral) dc:dentifier Assay dantifier (URI) rd:dstall Pathway name (iteral) dc:dentifier Assay dantifier (URI) rd:dstabel Pathway name (iteral)		NC	dc:identifier	HGNC identifier (URI)		rez	dc:identifie	r	UniProt identifier (URI)			
a MESH-D001244 dc:dentifier Association identifier (URI) MESH-C0003047 MESH-D003047 MESH-C0003047 MESH-D002047 dc:dentifier Score (Iteral) dc:dentifier MI2047 dc:dentifier MI2047 dc:dentifier MI2047 dc:dentifier MI2047 dc:dentifier MI2047 dc:dentifier MI2047 dc:dentifier PREDICATE OBJECT PREDICATE OBJECT Vp:target Entrez gene identifier (URI) PREDICATE OBJECT Visticabil edam-data_1027 Entrez gene identifier (URI) Prechicate OBJECT redicatinifier Assay utentifier (URI) Prechicate OBJECT redicatinifier Resay utentifier (URI) Prechicate OBJECT redictiontifier <t< td=""><th></th><th>HG</th><td>edam:data_2298</td><td>HGNC identifier (literal)</td><td></td><th>Ent</th><td colspan="2" rowspan="2"></td><td colspan="2" rowspan="2"></td><td></td></t<>		HG	edam:data_2298	HGNC identifier (literal)		Ent						
get_identifier Association identifier (URI) MESH-C0003047 MESH Disease identifier (URI) dc:identifier MESH-Disease identifier (URI) dc:identifier Source (literal) admidata_1027 Entrez gene identifier (URI) mutotare Protein source (literal) admidata_1027 Entrez gene identifier (URI) mutotare Protein source (literal) mutotare Protein source (literal) mutotare Protein source (literal) mutotare Protein source (literal) admidata_1027 Entrez gene identifier (URI) mutotare Protein source (literal) admidata_1027 Entrez gene identifier (URI) motitaxon:131667 Taxonomy (literal) a MMX:000041 dc:source Source (literal) motitaxon:131667 Taxonomy (URI) dc:source Source (literal)			skos:exactMatch	Matched identifier (URI)							-	
generation dc:identifier Association identifier (URI) MESH:C0003047 MESH Disease identifier (URI) dc:identifier Source (literal) edam.data_1027 Entrez gene identifier (URI) a Mi2047 ungoted Protein source (literal) edam.data_1027 Entrez gene identifier (URI) protein source (literal) Entrez gene identifier (URI) ungoted Protein source (literal) edam.data_1027 Entrez gene identifier (URI) protein source (literal) Entrez gene identifier (URI) torize gene identifier (URI) PREDICATE OBJECT precticates entrez gene identifier (URI) PREDICATE OBJECT interaze (literal) Entrez gene identifier (URI) PREDICATE OBJECT interaze (literal) Entrez gene identifier (URI) PREDICATE OBJECT interaze (literal) Entrez gene identifier (URI) Prectoicates OBJECT interaze (literal) Prectoicates OBJECT Figure 1: Resource Descripter indicatentifier Pathway identifier (URI) Framework schema of AOP-V AOP-V inditaton:131567 Taxonomy (URI) <th colspan="12"></th>												
geogram dcsource Source (ilteral) dcs.dentifier Entrez gene identifier (URI) a ML2047 a ML2047 dcs.identifier Interaction ID (URI) wp:source Protein source (itteral), Entrez gene identifier (URI) vp.target Protein target (iteral), Entrez gene identifier (URI) a PW000001 dc:identifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway name (iteral) edam.data_1027 Entrez gene identifier (URI) nobitaxon:131567 Taxonomy (URI) dc:identifier Assay identifier (URI) dc:identifier Assay identifier (URI) edam.data_1027 Entrez gene identifier (URI) nobitaxon:131567 Taxonomy (URI) dc:identifier Assay identifier (URI) rd:iabel Assay identifier (URI) rd:iabel Assay name (iteral) rd:iabel Assay name (iteral) rd:iabel Assay name (iteral) rd:iabel Assay name (iteral)	Discose-Gono		а	MESH:D001244			а	1		MESH:C0003047		
geogram dc:source Source (iteral) dc:source Source (iteral) edam.data_1027 Entrez gene identifier (URI) a Mi2047 geogram Entrez gene identifier (URI) wp:target Protein source (iteral) condentifier Protein source (iteral) motions.MTHU024701 Score (iteral) a Mi2047 wp:target Protein source (iteral) condentifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway name (iteral) dc:identifier Assay identifier (URI) dc:identifier Assay name (iteral) rd:identifier Assay name (iteral) rd:idatel Assay name (iteral) rd:identifier Assay name (iteral) rd:identifier		u a	dc:identifier	Association identifier (URI)		ase	dc:identifier		MESH Disease identifier (URI)			
geogram dcsource Source (ilteral) dcs.dentifier Entrez gene identifier (URI) a ML2047 a ML2047 dcs.identifier Interaction ID (URI) wp:source Protein source (itteral), Entrez gene identifier (URI) vp.target Protein target (iteral), Entrez gene identifier (URI) a PW000001 dc:identifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway identifier (URI) dc:identifier Pathway name (iteral) edam.data_1027 Entrez gene identifier (URI) nobitaxon:131567 Taxonomy (URI) dc:identifier Assay identifier (URI) dc:identifier Assay identifier (URI) edam.data_1027 Entrez gene identifier (URI) nobitaxon:131567 Taxonomy (URI) dc:identifier Assay identifier (URI) rd:iabel Assay identifier (URI) rd:iabel Assay name (iteral) rd:iabel Assay name (iteral) rd:iabel Assay name (iteral) rd:iabel Assay name (iteral)		iatic	MESH:C0003047	MESH Disease identifier (URI)	╶╶╎╴┤╼╸	Dise	rdfs:label		Disease Name (literal)			
edam:data_1027 Entrez gene identifier (URI) a Mi:2047 dc:dontififor Interaction ID (URI) wp:source Protein source (iteral), Entrez gene identifier (URI) wp:target Protein target (iteral), Entrez gene identifier (URI) LOINC:MTHU024701 Score (iteral) a PW:000001 dc:identifier Pathway identifier (URI) rds:label Pathway name (iteral) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (iteral) a MMC:0000411 dc:source Assay source name (iteral) rds:label Assay source name (iteral) ncbitaxon:131567 Taxonomy (URI) dc:source Assay source name (iteral) rds:label Assay source name (iteral) ncbitaxon:131567 Taxonomy (URI) dc:source Assay source name (iteral) rds:label Assay source name (iteral) ncbitaxon:131567 Taxonomy (URI) UBERON.0000061 Tissue name (iteral)		SOC	dc:source	Source (literal)		-	dcterms:isl	PartOf	Disease-Gene Association identifier			
A MI2047 dc:dentifier htteraction ID (UR) wp:source Protein source (iteral), Entrez gene identifier (UR) wp:source Pretoin target (iteral), Entrez gene identifier (UR) wp:source A motoconoon A c:dentifier A c:source A c:dentifier A c:dentifier A c:source A c:dentifier A c:source A c:dentifier A c:source A c:source A c:source A c:dentifier A c:source A c:source A c:dentifier A c:source A c:s		As	LOINC:MTHU024701	Score (literal)								
de:identifier Interaction ID (URI) wp:source Protein source (Iteral) wp:target Protein target (Iteral) LOINC:MTHU024701 Score (Iteral) a PW:000001 dc:identifier Pathway identifier (URI) rdfs:label Pathway identifier (URI) edam:data_1027 Entrez gene identifier (URI) rdfs:label Pathway name (Iteral) edam:data_1027 Entrez gene identifier (URI) nobitaxon:131567 Taxonomy (URI) dc:identifier Assay identifier (URI) rdfs:label			edam:data_1027	Entrez gene identifier (URI)								
de:identifier Interaction 10 (URI) wp:source Protein source (Iteral). wp:target Protein target (Iteral). LOINC:MTHU024701 Score (Iteral) a PW:000001 dc:identifier Pathway identifier (URI) rdfs:label Pathway name (Iteral) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (Iteral) a MMO:000041 dc:source Assay identifier (URI) rdfs:label Assay identifier (URI) <th rowspan="5">Protein-Protein</th> <th></th> <th>а</th> <th colspan="7">MI:2047</th> <th></th>	Protein-Protein		а	MI:2047								
LOINC:MTHU024701 Score (literal) a PW:000001 dc:dentifier Pathway identifier (URI) rdfs:label Pathway name (literal) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MM0:0000441 dc:dentifier Assay identifier (URI) rdfs:label Assay source name (literal) rdfs:label Assay source name (literal) rdfs:label Assay name (literal) rdfs:label Assay name (literal) rdfs:label Assay name (literal) rcbitaxon:131567 Taxonomy (URI) DEECNV:000061 Tissue name (literal)		u u	dc:identifier	Interaction ID (URI)				2090	/			
LOINC:MTHU024701 Score (ilteral) a PW:0000001 dc:dentifier Pathway identifier (URI) rdfs:label Pathway name (ilteral) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (ilteral) a MMO:0000411 dc:source Assay identifier (URI) dc:source Assay source name (ilteral) rdfs:label Assay source name (ilteral) rdfs:label Assay name (ilteral) <		in-rro	wp:source	Protein source (literal), Entrez gene identifier (URI)			H	PREDICAT	Έ	OBJECT	1	
dc:dentifier Pathway identifier (URI) rdfs:label Pathway name (literal) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MMO:0000441 dc:identifier Assay identifier (URI) dc:identifier Assay identifier (URI) dc:identifier Assay ource name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) uccource Assay source name (literal) ncbitaxon:131567 Taxonomy (URI) uBERON:0000061 Tissue name (literal)		Inte	wp:target	Protein target (literal), Entrez gene identifier (URI)								
dc:dentifier Pathway identifier (URI) rdfs:label Pathway name (literal) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MMO:0000441 dc:dentifier Assay identifier (URI) dc:dentifier Assay identifier (URI) dc:source Assay identifier (URI) dc:dentifier Assay identifier (URI) dc:source Assay source name (literal) ncbitaxon:131567 Taxonomy (URI) uberrow Assay name (literal) ncbitaxon:131567 Taxonomy (URI) uberrow UBERON:0000061 Tissue name (literal) Tissue name (literal)			LOINC:MTHU024701	Score (literal)			Ъ С	PREDICAL	E	OBJECT		
dc:dentifier Pathway identifier (URI) rdfs:label Pathway name (literal) edam:data_1027 Entrez gene identifier (URI) ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MMO:0000441 dc:dentifier Assay identifier (URI) dc:source Assay source name (literal) ncbitaxon:131567 Taxonomy (URI) uBERON:0000061 Tissue name (literal)			а	PW:0000001			D.		F			
ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MMO:0000441 dc:dentifier Assay identifier (URI) dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:0000061 Tissue name (literal)		_	dc:identifier	Pathway identifier (URI)			0)	FILDICAT	L	OBJECT		
ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MMO:0000441 dc:dentifier Assay identifier (URI) dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:0000061 Tissue name (literal)		(ma)	rdfs:label	Pathway name (literal)								
ncbitaxon:131567 Taxonomy (URI) dc:source Source (literal) a MMO:0000441 dc:dentifier Assay identifier (URI) dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:0000061 Tissue name (literal)		path	edam:data_1027	Entrez gene identifier (URI)						-		
a MMO:0000441 dc:identifier Assay identifier (URI) dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:000061 Tissue name (literal)		-	ncbitaxon:131567	Taxonomy (URI)		Framework schema of AOP-Wike extended with AOP-DB. All subjects						
dc:identifier Assay identifier (URI) dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:0000061 Tissue name (literal)			dc:source	Source (literal)								
dc:identifier Assay identifier (URI) dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:0000061 Tissue name (literal)			а	MMO:0000441						-		
dc:source Assay source name (literal) rdfs:label Assay name (literal) ncbitaxon:131567 Taxonomy (URI) UBERON:0000061 Tissue name (literal)		oxcast Assay	dc:identifier	Assay identifier (URI)		types) are highly connective in the RDF						
			dc:source	Assay source name (literal)								
			rdfs:label	Assay name (literal)		Shades of green indicate AOP-Wiki,						
			ncbitaxon:131567	Taxonomy (URI)							KI,	
bao:BAO_0003064 Entrez gene identifier (URI)		Ĕ	UBERON:0000061	Tissue name (literal)		shades of blue indicate AOP-DB.						
			bao:BAO_0003064	Entrez gene identifier (URI)								

[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.