

Open camera or QR reader and
scan code to access this article
and other resources online.



Providing Adverse Outcome Pathways from the AOP-Wiki in a Semantic Web Format to Increase Usability and Accessibility of the Content

Marvin Martens,¹ Chris T. Evelo,^{1,2} and Egon L. Willighagen¹

Abstract

Introduction: The AOP-Wiki is the main platform for the development and storage of adverse outcome pathways (AOPs). These AOPs describe mechanistic information about toxicodynamic processes and can be used to develop effective risk assessment strategies. However, it is challenging to automatically and systematically parse, filter, and use its contents. We explored solutions to better structure the AOP-Wiki content, and to link it with chemical and biological resources. Together, this allows more detailed exploration, which can be automated.

Materials and Methods: We converted the complete AOP-Wiki content into resource description framework (RDF) triples. We used >20 ontologies for the semantic annotation of property–object relations, including the Chemical Information Ontology, Dublin Core, and the AOP Ontology.

Results: The resulting RDF contains >122,000 triples describing 158 unique properties of >15,000 unique subjects. Furthermore, >3500 link-outs were added to 12 chemical databases, and >7500 link-outs to 4 gene and protein databases. The AOP-Wiki RDF has been made available at <https://aopwiki.rdf.bigcat-bioinformatics.org>

Discussion: SPARQL queries can be used to answer biological and toxicological questions, such as listing measurement methods for all Key Events leading to an Adverse Outcome of interest. The full power that the use of this new resource provides becomes apparent when combining the content with external databases using federated queries.

Conclusion: Overall, the AOP-Wiki RDF allows new ways to explore the rapidly growing AOP knowledge and makes the integration of this database in automated workflows possible, making the AOP-Wiki more FAIR.

Keywords: adverse outcome pathways, linked data, resource description framework, risk assessment, SPARQL

Introduction

SINCE ITS ESTABLISHMENT in 2010, the adverse outcome pathway (AOP) concept has become a prominent tool for the risk assessment community.^{1,2} AOPs are a chain of

biological processes, called Key Events (KEs), starting from a molecular perturbation with a stressor toward an Adverse Outcome (AO), connected by Key Event Relationships (KERs). AOPs exist to capture all mechanistic toxicological knowledge from literature and data, to direct future

¹Department of Bioinformatics—BiGCaT, NUTRIM, and ²Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands.

A preprint of this work has been published on ChemRxiv (chemrxiv.org), doi:10.26434/chemrxiv.13524191

© Marvin Martens et al., 2022; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons License [CC-BY] (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

studies to fill gaps of existing knowledge, and to drive Integrated Approaches to Testing and Assessment (IATA) development.^{1,3} This was demonstrated with the AOP-based IATA for skin sensitization, resulting in various IATA with combinations of *in vitro* and *in silico* assays outperforming animal tests.⁴

The majority of the AOPs are developed and stored in the AOP-Wiki (<https://aopwiki.org/>), which is part of the AOP knowledge base, released in 2014 as a result of the AOP development program initiated by the Organisation for Economic and Collaborative Development.⁵ This wiki is designed to facilitate collaborative development of qualitative AOP descriptions, and thereby promote their incorporation into risk assessments and stimulate effective reuse of mechanistic toxicological knowledge.^{6,7}

The resulting AOPs describe much of the biological context surrounding toxicological processes; most of the information on genes, chemicals, biological pathways, phenotypes, among other things, is already captured in specialized databases or ontologies outside of AOP-Wiki.⁸ However, the AOP-Wiki has limited possibilities for linking external information and data, mostly consisting of free-text descriptions and links to the U.S. CompTox Chemistry Dashboard⁹ and to NCBI for taxonomic applicability.¹⁰ An initiative to make the reporting more consistent was the introduction of KEs Components¹¹ for the annotation of Biological Processes, Biological Objects, and Biological Actions for KEs, and annotations of cell types and organs in which KEs can occur.

Since the AOP-Wiki is the central repository for AOPs and therefore a key player in the shift toward animal-free testing strategies, it is essential that its contents can be queried and utilized effectively to answer biological questions and to reuse existing knowledge. However, accessing the data computationally or linking with other resources is hardly possible when only downloadable eXtensible Markup Language (XML) data dumps are provided that consist mostly of free text. Because of these aspects, parsing and querying the continuously growing amount of information in the AOP-Wiki is a complex, time-consuming task. This is a problem because it prevents the integration of AOP knowledge with other data and resources.

This could be resolved by applying Linked Open Data solutions, such as structuring the data in a resource description framework (RDF) model,¹² introducing persistent identifiers and semantic annotations, and implementing Application Programming Interfaces (APIs) for accessing the data. RDF represents knowledge as semantic triples, in which a subject, predicate, and object together define a statement and assist in the meaningful representation of knowledge in a machine-readable manner.

These concepts are generally in line with the FAIR principles¹³ for data and knowledge management, developed to enhance the Findability, Accessibility, Interoperability, and Reusability of data and allow computational support of data usage; for example, the solutions applied by the Swiss Institute of Bioinformatics with the development of neXtProt Linked Data by implementing RDF annotations for easier exploration and retrieval of data through web services.^{14,15}

Also, the use of ontologies and vocabularies for semantic annotations allows for the integration of data between resources, such as the direct linking of chemical or protein databases with WikiPathways.^{16,17}

In this article, we show how using RDF makes the AOP-Wiki content more usable for automated exploration in combination with other existing semantic web-based information sources. We describe our implementation of Linked Open Data solutions for the AOP-Wiki to introduce new, effective ways of accessing and using the data. These solutions will enhance the usefulness of the AOP-Wiki to risk assessors, developers, and modelers, and facilitate answering complex research questions, also across databases or as part of automated workflows.

We hypothesize that with the implementation of RDF, with the use of standard ontologies for semantic modeling of information captured in AOPs, the data can be better exploited.¹⁸ Furthermore, the use of persistent, unique, and resolvable identifiers allows interoperability with other related data sources. When combined with computational tools that can access experimental data, these approaches can make AOP information a core element for predictive modeling.¹⁹

Materials and Methods

Registering AOP-Wiki identifiers in Identifiers.org

Before the development of the AOP-Wiki RDF, we registered the identifiers for the AOP, KE, KER, and stressor in the Minimum Information Required In the Annotation of Models (MIRIAM) Registry²⁰ to allow Identifiers.org to resolve Internationalized Resource Identifiers (IRIs). To make all identifiers in the AOP-Wiki resolvable and linking to their corresponding database webpages, these IRIs, along with a variety of chemical and gene database identifier types, were implemented in the AOP-Wiki RDF.

XML-to-RDF conversion code

The code for the XML-to-RDF conversion was written as a Jupyter notebook using Python version 3.7.3 in JupyterLab version 0.35.5, and is stored in GitHub (<https://github.com/marvinm2/AOPWikiRDF>).²¹

Downloading and parsing the AOP-Wiki XML. It downloads the AOP-Wiki XML quarterly download file of January 1, 2021 from <https://aopwiki.org/downloads> and parses the file with the ElementTree XML API Python library. Next, the Jupyter notebook stores all the AOP-Wiki content in a Python nested dictionary data model, one for each of the main components, which form the basis of the existing AOP-Wiki. These are the AOPs, KEs, KERs, stressors, chemicals, taxonomy, cell-terms, organ-terms, and the KE components, which comprise Biological Processes, Biological Objects, and Biological Actions.

Semantic annotation in the RDF. Terms from common biomedical terminologies and standard metadata vocabularies were used as predicates. These terms were retrieved from BioPortal²² or in the corresponding Web Ontology Language (OWL)²³ files stored in GitHub. These ontologies include Dublin Core,²⁴ DCMI Metadata Terms,²⁵ RDF Schema,²⁶ Friend Of A Friend,²⁷ Adverse Outcome Pathway Ontology,²⁸ Phenotypic Quality Ontology,²⁹ Chemical Information ontology,³⁰ NCI Thesaurus,³¹ Measurement Method Ontology,³² Simple Knowledge Organization System,³³

TABLE 1. ONTOLOGIES AND VOCABULARIES USED IN THE RESOURCE DESCRIPTION FRAMEWORK

<i>Ontology name</i>	<i>Prefix in RDF</i>	<i>Base IRI</i>
Dublin Core ²⁴	dc	http://purl.org/dc/elements/1.1/
DCMI Metadata Terms ²⁵	dcterms	http://purl.org/dc/terms/
RDF Schema ²⁶	rdfs	www.w3.org/2000/01/rdf-schema#
Friend Of A Friend ²⁷	foaf	http://xmlns.com/foaf/0.1/
Adverse Outcome Pathway Ontology ²⁸	aopo	http://aopkb.org/aop_ontology#
Phenotypic Quality Ontology ²⁹	pato	http://purl.obolibrary.org/obo/PATO_
Chemical Information Ontology ³⁰	cheminf	http://semanticscience.org/resource/CHEMINF_
NCI Thesaurus ³¹	nci	http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#
Measurement Method Ontology ³²	mmo	http://purl.obolibrary.org/obo/MMO_
Simple Knowledge Organization System ³³	skos	www.w3.org/2004/02/skos/core#
National Center for Biotechnology Information Organismal Classification ¹⁰	ncbitaxon	http://purl.bioontology.org/ontology/NCBITAXON/
Gene Ontology ³⁴	go	http://purl.obolibrary.org/obo/GO_
EDAM bioinformatics operations, types of data, data formats, identifiers, and topics ³⁵	edam	http://edamontology.org/
Provenance, Authoring, and Versioning ³⁶	pav	http://purl.org/pav/
Vocabulary of Interlinked Datasets ³⁷	void	http://rdfs.org/ns/void#
Data Catalog Vocabulary ³⁸	dcat	http://www.w3.org/ns/dcat#

IRIs, Internationalized Resource Identifiers.

National Center for Biotechnology Information Organismal Classification,¹⁰ Gene Ontology,³⁴ EDAM bioinformatics operations, types of data, data formats, identifiers, and topics,³⁵ Provenance, Authoring and Versioning,³⁶ Vocabulary of Interlinked Datasets,³⁷ Data Catalog Vocabulary.³⁸

Table 1 provides an overview of these, including their prefixes and IRI patterns. Furthermore, the IRIs were completed for annotations that already exist in the AOP-Wiki such as the KE components, cell-terms, and organ-terms. These annotations include terms of the Cell Ontology,³⁹ Uber-anatomy ontology,⁴⁰ Gene Ontology,³⁴ Molecular Interactions Controlled Vocabulary,⁴¹ Mammalian Phenotype Ontology,⁴² Medical Subject Headings,⁴³ Human Phenotype Ontology,⁴⁴ Population and Community Ontology,^{45,46} Neuro Behavior Ontology,⁴⁷ Vertebrate trait ontology,⁴⁸ PRotein Ontology,⁴⁹ Chemical Entities of Biological Interest,⁵⁰ and Foundational Model of Anatomy Ontology.⁵¹ These ontologies are listed in Supplementary Table S1 (Annex) together with their prefixes and IRI patterns.

Addition of gene and protein identifiers. To increase the number of annotations and add more types of gene and protein identifiers for improved linking of data and repositories, the XML-to-RDF conversion includes two methods of mapping to gene and protein identifiers.

The first of which is based on existing Biological Object annotations with PRotein ontology (PR) terms⁴⁹ in the AOP-Wiki, which were mapped to identifiers from NCBI Gene⁵² and UniProt,⁵³ and symbols from the HUGO Gene Nomenclature Committee (HGNC)⁵⁴ with the PR mapping file, promapping.txt, downloaded from <https://proconsortium.org/download/current/> on May 10, 2020.

The second method involved textual gene identifier mapping for KEs and KERs, for which we extracted approved symbols, names, and synonyms for all human genes from the HGNC (downloaded from <https://www.genenames.org/>⁵⁴ in January 2020). After loading the HGNC file, a symbol dictionary was created, which included all official gene symbols, names, and alternative gene names, and the

library has been extended with variants of textual separators surrounding the symbols to avoid partial word overlaps. Next, text matching was done for each KE and KER description, the Molecular Initiating Event (MIE) and AO-specific section of KEs, and biological plausibility and empirical support sections of KERs. For each perfect match, the matching HGNC identifier was added to the KE and KER dictionaries to store KE-gene information.

BridgeDb identifier mapping. On top of the chemicals already present in the AOP-Wiki and the genes and proteins IDs added to the RDF with the textual mapping of HGNC symbols, we extended the coverage of external molecular databases using BridgeDb, an identifier mapping service for chemicals, genes, proteins, and interactions.⁵⁵

The “requests” Python library (version 2.22.0) was used for calling BridgeDb’s “xref” function to perform identifier mapping for chemicals and HGNC IDs that resulted from the textual mapping. The BridgeDb service was loaded with the Metabolite BridgeDb ID Mapping Database (version HMDB4.0.20190116-CHEBI193-WIKIDATA20201104, released on November 4, 2020)⁵⁶ and the Gene/Protein BridgeDb ID Mapping Database (version 91, released on May 9, 2018).⁵⁷ For chemicals, the CAS IDs from the AOP-Wiki XML were used as input to retrieve identifiers from ChEBI,⁵⁰ ChemSpider,⁵⁸ Wikidata,^{59,60} ChEMBL,⁶¹ PubChem,⁶² Drugbank,⁶³ KEGG,⁶⁴ LIPID MAPS,⁶⁵ and HMDB.⁶⁶ For genes, the HGNC IDs were used to request matching identifiers for NCBI Gene, UniProt, and Ensembl.⁶⁷

File creation. All AOP-Wiki content, persistent identifiers, ontology annotations, and additional information for chemicals, genes, and proteins were stored into three RDF files using Turtle (ttl) syntax. While the central AOP-Wiki RDF file (AOP-Wiki.ttl) contains all existing AOP-Wiki components plus added chemical identifiers and identifiers mapped from PR terms in Biological Objects, the second file (AOP-Wiki-genes.ttl) contains all the text-mapped

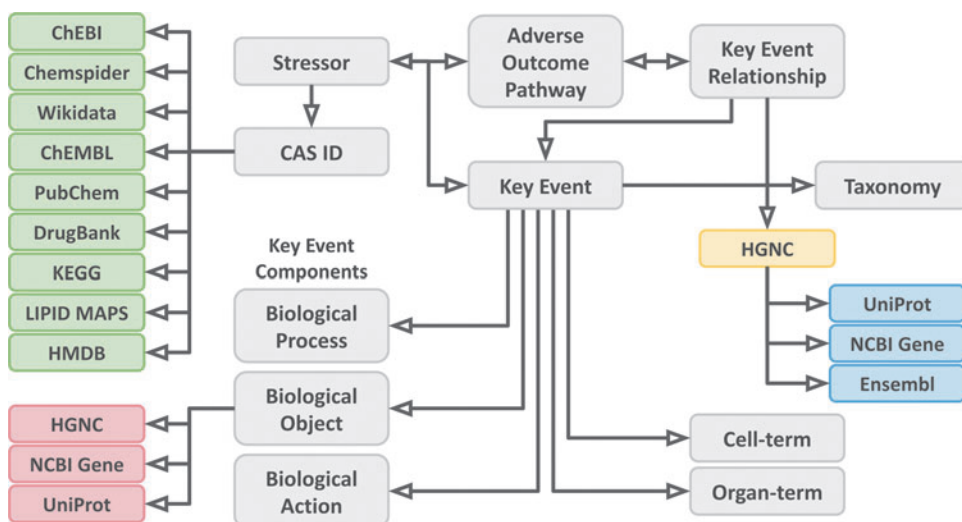


FIG. 1. General overview of the AOP-Wiki RDF scheme. Arrows show the directional relationships described in the RDF. Gray boxes represent the basic elements of the AOP-Wiki. Green boxes indicate added chemical IDs using BridgeDb. Red boxes indicate added gene/protein IDs using Protein Ontology mapping. 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. RDF, resource description framework. Color images are available online.

gene IDs and matching identifiers added by BridgeDb (Fig. 1). These files are accompanied by a metadata file (AOP-Wiki-void.ttl), which describes the datasets, code, and provenance, using standard vocabularies for semantic annotations of metadata, such as Dublin Core,^{24,25} Data Catalog Vocabulary,³⁸ Friend of a Friend,²⁷ and Vocabulary of Interlinked Datasets.³⁷

Validation and testing

Validation of the RDF. The RDF files were validated with the IDLab Turtle validator,⁶⁸ an open-source RDF validator for Turtle syntax and XSD datatype errors.

Loading and testing the RDF. After validation of the RDF, the AOP-Wiki RDF was loaded in a public SPARQL endpoint (<https://aopwiki.rdf.bigcat-bioinformatics.org/sparql>) and is accessible through the developed SNORQL User Interface (<https://aopwiki.rdf.bigcat-bioinformatics.org/>, Fig. 2).

The data were tested with a Jupyter notebook that executes SPARQL queries through the SPARQL endpoint. These SPARQL queries retrieve metadata and the statistics for types of subjects, frequency of ontology usage, and the number of link-outs to the various databases.²¹ All SPARQL queries used for the testing of the RDF are available in the SPARQL examples panel in the AOP-Wiki SNORQL User Interface.

Validation of the SPARQL endpoint. On January 17, 2021, the AOP-Wiki SPARQL endpoint was registered in YummyData,⁶⁹ which monitors compliance with Linked Data standards and scoring each SPARQL endpoint on availability, freshness, operation, usefulness, validity, and performance to calculate the Umaka Score. The scoring is done on a daily basis by performing SPARQL queries and HTTP requests related to the various measures for each aspect. YummyData also provides feedback on how to improve the score, which was used for improving the AOP-Wiki RDF and SPARQL endpoint.

Results

The main result of this project is an RDF schema and scripts that lead to the production of RDF content for all AOP-Wiki content with additional semantic annotations, persistent identifiers, and extended identifiers for genes and chemicals, and consists of 122,576 unique triples consisting of 15,132 unique subjects, 158 unique predicates, and 53,087 unique objects (Fig. 1). The semantic annotation was done using 8 standard metadata vocabularies and 17 domain-specific ontologies and vocabularies. We here detail these results.

Adverse outcome pathways

The 316 AOP subjects have 26 different types of predicates to create triples (Table 2). The overall most used vocabulary for predicates is Dublin Core and its extended set of terms, creating triples for the identifier, title, alternative title, creator, abstract, description, source, access rights, creation, and modification date for AOPs. The majority of these predicates also exist for other subjects. The AOP Ontology (AOP) is used to connect AOPs to KE subjects with the predicates “has_key_event,” “has_molecular_initiating_event,” and “has_adverse_outcome,” and connect with KER subjects with the predicate “has_key_event_relationship.”

Other terms of the AOP were used for describing the overall applicability, life stage applicability, and weight of evidence. Furthermore, AOP subjects are unique to contain information on the quantitative considerations of the AOP, which is linked with “edam:operation 3799.” The link with stressors, overall assessment description, KE essentiality, and the potential applications of the AOP were annotated using the NCI Thesaurus terms “nci:C54571,” “nci:C25217,” “nci:C48192,” and “nci:C25725,” respectively. Finally, the sex applicability of AOPs is annotated with “pato:0000047,” which stands for biological sex.

KEs and KERs

While the majority of triples for KEs and KERs have predicates identical to the ones for AOPs, there are properties that are unique to the 1131 KEs and 1363 KERs (Table 3). For

AOP-Wiki SPARQL Endpoint: <https://aopwiki.rdf.bigcat-bioinformatics.org/sparql/>

SPARQL Query:

```

1 #This query takes all Stressor chemicals in the AOP-Wiki and their mappings to ChEBI IDs
2 #Then, these ChEBI IDs are used to look up human molecular pathways in WikiPathways
3
4 PREFIX wp: <http://vocabularies.wikipathways.org/wp#>
5
6 SELECT DISTINCT ?ChemicalName ?mappedid ?LinkedStressor (group_concat(?LinkedAOP;separator=
7 (STR(?PathwayTitle) AS ?PathwayName) ?PathwayURI
8
9 WHERE{
10 ?cheLook a cheminf:000000; dc:title ?ChemicalName ; dcterms:isPartOf ?LinkedStressor ;
11 skos:exactMatch ?mappedid .
12 ?LinkedStressor dcterms:isPartOf ?LinkedAOPURI .
13 ?LinkedAOPURI a aopo:AdverseOutcomePathway ; rdfs:label ?LinkedAOP.
14 ?mappedid a cheminf:000407; cheminf:000407 ?CHEBI.
15 SERVICE <http://sparql.wikipathways.org/sparql>{
16 ?metabolite wp:bdbChEBI ?mappedid ; dcterms:isPartOf ?PathwayURI.
17 ?PathwayURI a wp:Pathway ; dcterms:identifier ?PathwayID ;
18 dc:title ?PathwayTitle ; wp:organismName "Homo sapiens"^^xsd:string .}}
19 ORDER BY DESC (?ChemicalName)

```

SPARQL Examples:

<https://github.com/marvinm2/AOPWikiSNC>

Type part of the query file name to search for...

Search Clear

- A. Metadata
- B. Datadump
- C. Search
- D. Simpleconversions
- E. Chemical-centered
- F. Federated

1. PathwayswithChem.rq

SPARQL results (211 results)

ChemicalName	mappedid	LinkedStressor	AOPs	Pathway_ID	PathwayName	PathwayURI
tert-Butyl hydroperoxide	https://identifiers.org/chebi/CHEBI:64090	https://identifiers.org/aop.stressor/457	AOP 296	WP4008	NO/cGMP/PKG mediated neuroprotection	https://identifiers.org
Valproic acid	https://identifiers.org/chebi/CHEBI:39867	https://identifiers.org/aop.stressor/401	AOP 212 AOP	WP3871	Valproic acid pathway	https://identifiers.org

Powered by AOP-Wiki and AOP-Wiki SNORQL - Cookie Policy - **BiGCaT** See also Martens et al. (2021)

FIG. 2. The AOP-Wiki SNORQL user interface. The AOP-Wiki SNORQL user interface allows for user-friendly access to the AOP-Wiki RDF by syntax highlighting and through providing a SPARQL examples panel (right panel). Color images are available online.

KEs, these properties include measurement methods, level of biological organization, and structured information on cell-terms, organ-terms, Biological Processes, Objects, and Actions (Table 3). The measurement methods are coupled to 350 KEs with the predicate “mmo:0000000” from the Measurement Method Ontology, which stands for measurement method, and level of biological organization is linked with the “nci:C25664” to all KEs. Cell-terms and Organ-terms are described with the AOPO terms “aopo:CellType-Context” and “aopo:OrganCotext,” respectively.

The Biological Process triples have the predicate “go:008150,” which stands for biological process, and the Biological Objects and Actions are linked with “pato:0001241” and “pato:0000001,” respectively. These ontological annotations are connected through IRIs of other subjects in the RDF. A shared predicate with AOPs is the term “nci:C54571” for MIEs that have links to stressors in the AOP-Wiki.

Properties that are specific to the 1363 KERs are the biological plausibility, empirical support, uncertainties, which we linked with the predicates “nci:C80263,” “edam:data_2042,” and “nci:71478.” These stand for the rationale, evidence, and uncertainty, respectively. Also, the RDF con-

nects the upstream and downstream KEs of KERs with the terms “aopo:has_upstream_key_event” and “aopo:has_downstream_key_event” from the AOPO (Table 3).

Stressors and chemicals

The RDF contains general stressor information such as descriptions and identifiers, and stressor triples describe their connections to MIEs and AOPs with “dcterms:isPartOf.” Sixty-three percent of the 523 stressors are also linked to chemicals with the predicate “aopo:has_chemical_entity” (Table 4).

The 329 chemical subjects are annotated with CAS and CompTox identifiers, and 320 also have InChIKeys, all of which we annotated with the Chemical Information Ontology (Table 4). Furthermore, the chemicals have predicate “skos:exactMatch” to link to all mapped chemical subjects present in the RDF, providing link-outs to nine additional external databases (Table 5). We annotated these with “rdf:type” and terms from the Chemical Information Ontology. In total, there are 3904 link-outs to 12 different chemical databases, allowing users to explore the AOP-Wiki by using their preferred type of chemical identifiers.

TABLE 2. ADVERSE OUTCOME PATHWAYS AND THEIR PROPERTIES IN RESOURCE DESCRIPTION FRAMEWORK

<i>Predicate</i>	<i>Object</i>	<i>Object example</i>
Adverse outcome pathway	aopo:AdverseOutcomePathway	
a	Adverse Outcome Pathway (IRI)	aop:38
dc:identifier	Label (literal)	“AOP 38”
rdfs:label	Title (literal)	“Protein Alkylation leading to Liver Fibrosis”
dc:title	Alternative title (literal)	“Protein Alkylation to Liver Fibrosis”
dcterms:alternative	Author (literal)	““““Brigitte Landesmann...
dc:creator	Abstract (literal)	““““Hepatotoxicity in general is of special interest... aop.stressor:9,aop.stressor: 13,aop.stressor:60,... aop.events:55,aop.events:1492,aop.events:1493, ... aop.events:244
dcterms:abstract	Stressor (IRI)*	
nci:C54571	Key Event (IRI)*	
aopo:has_key_event	Key Event (IRI)*	
aopo:has_molecular_initiating_event	Key Event (IRI)*	aop.events:344
aopo:has_adverse_outcome	Key Event Relationship (IRI)*	aop.relationships:269,aop.relationships:1718, ...
aopo:has_key_event_relationship	Description (literal)	““““Two prototypical chemicals acting via protein alkylation are ... “Unspecific” “Not Otherwise Specified”
dc:description	Sex applicability (literal)	““““The described AOP is valid for both sexes and any life stage ...
pato:0000047	Life stage applicability (literal)	““““More advanced in vitro models systems are needed ...
aopo:LifeStageContext	Applicability (literal)	““““Support for Essentiality of KEs ... ““““This systematic and coherent display of currently available ... ““““Assessment of the Weight-of-Evidence supporting the AOP ... ““““The essentiality of each of the KEs for this AOP ...
aopo:AopContext	Quantitative considerations (literal)	“Open for citation & comment” <https://identifiers.org/aop/38> “2016-11-29T18:41:16” “2019-04-30T12:53:51” “AOPWiki”
edam:operation_3799	Weight of evidence (literal)	
aopo:has_evidence	Potential applications (literal)	
nci:C25725	Overall assessment (literal)	
nci:C25217	Key Event essentiality (literal)	
nci:C48192	AOP status (literal)	
dc:accessRights	Webpage (URL)	
foaf:page	Date of creation (literal)	
dcterms:created	Date of latest modification (literal)	
dcterms:modified	Source (literal)	
dc:source		

Asterisks indicate the object IRIs that connect to other subjects in the RDF.

Ontological annotations

Since the taxonomies, cell-terms, organ-terms, and the KE components all already have ontological annotations in the AOP-Wiki, they have the same properties that describe their type (Table 6), identifier, title, and source. These titles are based on the user-provided entries in the AOP-Wiki. Unique for the biological objects annotated with the PR is the inclusion of the “skos:exactMatch” predicate linking to 576 matching identifiers from UniProt, HGNC, and NCBI Gene (Table 7) to 126 PR tags, which are used in 166 KEs.

Gene and protein identifiers

Extending the links of KEs and KERs with genes and proteins, RDF triples of 846 unique text-mapped gene identifiers on KEs and KERs are stored in a separate file. These make triples of KE and KER subjects link to the mapped HGNC identifiers with the predicate “edam:data_1025,” which stands for Gene identifier. These HGNC identifiers are subjects themselves, and have the “skos:exactMatch” predicate to link to matching identifiers from UniProt, NCBI Gene, and Ensembl, providing a total of 6001 link-outs using this method (Table 7).

Federated SPARQL query example

The addition of external identifiers facilitates the execution of federated SPARQL queries to combine resources, such as the example shown in Figure 2, which is located in the SPARQL example panel under the folder “F. Federated.” The SPARQL query looks up all entities defined as a Chemical in the AOP-Wiki RDF with the predicate and subject “a cheminf:000000,” and extracts their names, mapped identifiers, and linked stressors. Next, the query looks for AOPs that mention the stressor. The next line in the query restricts the results by explicitly adding the type of “a aopo:AdverseOutcomePathway.” Similarly, the query makes sure that the mapped identifier of the chemical is from the ChEBI database by defining the mapped identifier as type “cheminf:000407.”

This is followed by the “SERVICE” statement, which defines the external SPARQL endpoint and defines the federated part of the query. In this example, the external SPARQL endpoint is WikiPathways (<https://sparql.wikipathways.org/sparql>), where the ChEBI identifier is used to match relevant pathways using “dcterms:isPartOf.” The last part of the SPARQL query retrieves information about the pathway and filters for human pathways. This

TABLE 3. KEY EVENTS AND THEIR PROPERTIES IN RESOURCE DESCRIPTION FRAMEWORK

<i>Predicate</i>	<i>Object</i>	<i>Object example</i>
Key event		
a	aopo:KeyEvent	
dc:identifier	Key Event (IRI)	aop.events:1502
rdfs:label	Label (literal)	“KE 1502”
dc:title	Title (literal)	“Histone deacetylase inhibition”
dcterms:alternative	Alternative title (literal)	“Histone deacetylase inhibition”
nci:C25664	Level of biological organization (literal)	““““Molecular””””
dc:description	Description (literal)	““““The inhibition of HDAC by HDIs is well conserved... hgnc:HDAC9,hgnc:MAA,hgnc:PRDX2 ““““The measurement of HDAC inhibition monitors changes... aop.stressor:340,aop.stressor:341,aop.stressor:342,... cl:0000000 uberon:0000062 go:0004857 pr:000008478 “WIKI:2” ncbitaxon:10116,”WCS_9606”,ncbitaxon:10090 “Unspecific” “All life stages”
edam:data_1025	HGNC ID (IRI)*	
mmo:0000000	Measurement method (literal)	
nci:C54571	Stressor (IRI)*	
aopo:CellTypeContext	Cell-term (IRI)*	
aopo:OrganContext	Organ-term (IRI)*	
go:0008150	Biological process (IRI)*	
pato:0001241	Biological object (IRI)*	
pato:0000001	Biological action (IRI)*	
ncbitaxon:131567	Taxonomy (IRI or literal)*	
pato:0000047	Sex applicability (literal)	
aopo:LifeStageContext	Life stage applicability (literal)	
foaf:page	Webpage (URL)	<https://identifiers.org/aop.events/1502>
dcterms:isPartOf	Adverse Outcome Pathway (IRI)*	aop:212,aop:274,aop:275
dc:source	Source (literal)	“AOPWiki”
Key event relationship		
a	aopo:KeyEventRelationship	
dc:identifier	Key Event Relationship (URI)	aop.relationships:865
rdfs:label	Label (literal)	“KER 865”
aopo:has_upstream_key_event	Key Event (URI)*	aop.events:844
aopo:has_downstream_key_event	Key Event (URI)*	aop.events:845
dc:description	Description (literal)	““““One of the oxidation products of uroporphyrinogen... ““““Reduced UROD enzyme activity, not protein levels... ““““Include consideration of temporal concordance... ““““The precise mechanism of UROD inhibition has yet... hgnc:UROD ncbitaxon: 10090,ncbitaxon:10116,“WCS_9606” “Unspecific” “All life stages”,“Adult”,“Juvenile”
nci:C80263	Biological plausibility (literal)	
edam:data_2042	Empirical support (literal)	
nci:C71478	Uncertainties or inconsistencies (literal)	
edam:data_1025	HGNC ID (IRI)*	
ncbitaxon:131567	Taxonomy (IRI or literal)*	
pato:0000047	Sex applicability (literal)	
aopo:LifeStageContext	Life stage applicability (literal)	
foaf:page	Webpage (URL)	<https://identifiers.org/aop.relationships/865>
dcterms:created	Date of creation (literal)	2016-11-29T18:41:35”
dcterms:modified	Date of latest modification (literal)	“2018-05-30T10:58:18”
dcterms:isPartOf	Adverse Outcome Pathway (URI)*	aop:131

Asterisks indicate the object IRIs that connect to other subjects in the RDF.

SPARQL query results in a table of 211 rows with each chemical, their ChEBI identifier, related stressor and AOP, and the molecular pathways that the chemical is involved in. These linked pathways provide additional insights into the general functions of chemicals and their involvement in cellular processes and responses.

Validation by YummyData

As an external validation of the developed SPARQL endpoint and RDF, YummyData indexes and ranks the AOP-Wiki SPARQL endpoint based on an array of tests on a daily basis (<https://yummydata.org/endpoint/142>). As of

TABLE 4. STRESSORS AND CHEMICALS AND THEIR PROPERTIES IN RESOURCE DESCRIPTION FRAMEWORK

<i>Predicate</i>	<i>Object</i>	<i>Object example</i>
Stressor		
a	nci:C54571	
dc:identifier	Stressor (IRI)	aop.stressor:208
rdfs:label	Label (literal)	“Stressor 208”
dc:title	Title (literal)	“Gemfibrozil”
dc:description	Description (literal)	““““Fibrate drug””””
aopo:has_chemical_entity	Chemical identifier (IRI)*	cas:25812-30-0
foaf:page	Webpage (URL)	<https://identifiers.org/aop.stressor/208>
dcterms:created	Date of creation (literal)	“2016-11-29T18:42:27”
dcterms:modified	Date of latest modification (literal)	“2020-03-31T10:24:40”
dcterms:isPartOf	Adverse Outcome Pathway (IRI)* Key Event (IRI)*	aop.events:227,aop.events:1170,aop:18, aop:37,aop:51,aop:61,aop:323, ...
Chemical		
a	cheminf:000000	
	cheminf:000446	
dc:identifier	CAS identifier (IRI)	cas: 103-90-2
cheminf:000446	CAS identifier (literal)	“103-90-2”
dc:title	Title (literal)	“Acetaminophen”
dcterms:alternative	Synonyms (literal)	“4-Acetamidophenol”, “Paracetamol”, ...
cheminf:000059	InChIKey (IRI)	inchikey:RZVAJINKPMORJF-UHFFFAOYSA-N
cheminf:000568	CompTox identifier (IRI)	comptox:DTXSID2020006
skos:exactMatch	Matched identifier (IRI)*	chebi:46195,chemspider:1906,wikidata:Q57055, ...
dcterms:isPartOf	Stressor (IRI)*	aop.stressor:57

Asterisks indicate the object IRIs that connect to other subjects in the RDF.

February 2021, the AOP-Wiki SPARQL endpoint is considered A rank with a Umaka score >80, consistently scoring above average on all aspects. Incidentally, the Umaka score drops slightly <80, giving it a B rank.

Discussion

The work described in this article has led to the creation of AOP-Wiki RDF based on the existing AOP-Wiki XML, combined with a variety of ontologies and enriched with persistent identifiers. Besides, the data are extended with additional identifiers for chemicals, proteins, and genes. The RDF has been validated, loaded on a SPARQL endpoint, and tested using a Jupyter notebook, and is indexed in YummyData for external validation of the SPARQL endpoint. These developments made AOP-Wiki content ready for use in risk assessment workflows, through coding environments, or in federated SPARQL queries.

TABLE 5. GENE AND PROTEIN DATABASES, THEIR TYPE ANNOTATION, AND THE NUMBER OF IDENTIFIERS PRESENT IN THE RESOURCE DESCRIPTION FRAMEWORK

<i>Database</i>	<i>rdfs:type</i>	<i>No. 1</i>	<i>No. 2</i>
HGNC	edam:data_2298	97	846
Ensembl	edam:data_1033	—	813
Entrez Gene	edam:data_1027	32	804
UniProt	edam:data_2291	447	3653

Values in no. 1 are based on Protein Ontology mappings, and values in no. 2 are based on textual mapping with HGNC symbols. HGNC, HUGO Gene Nomenclature Committee.

Because the AOPO is developed for consistent reporting in the domain of AOPs and allowing the integration of data and tools,²⁸ it was an obvious choice to implement the AOPO for semantic annotations in the AOP-Wiki RDF. The ontology includes a variety of AOP-specific definitions for properties and classes, which are directly applicable to AOP-Wiki content in the RDF. However, these do not fully cover all types of entities and relationships that exist in the AOP-Wiki. For example, while it has terms to describe the connections between AOPs, KEs, and KERs, there is no annotation for the link with stressors. Similarly, terms are lacking for sex and taxonomic applicability, KE

TABLE 6. CHEMICAL DATABASES, THEIR TYPE ANNOTATION, AND THE NUMBER OF IDENTIFIERS PRESENT IN THE RESOURCE DESCRIPTION FRAMEWORK

<i>Database</i>	<i>rdfs:type</i>	<i>No.</i>
CAS	cheminf:000446	329
ChEBI	cheminf:000407	803
ChemSpider	cheminf:000405	343
ChEMBL compound	cheminf:000412	287
CompTox	cheminf:000568	329
Drugbank	cheminf:000406	161
HMDB	cheminf:000408	363
InChIKey	cheminf:000059	320
KEGG compound	cheminf:000409	264
Lipid maps	cheminf:000564	30
PubChem compound	cheminf:000140	338
Wikidata	cheminf:000567	328

TABLE 7. CELL-TERMS, ORGAN-TERMS, TAXONOMIES, KEY EVENT COMPONENTS, AND TYPE ANNOTATION IN THE RESOURCE DESCRIPTION FRAMEWORK

<i>Subject</i>	<i>rdfs:type</i>
Cell-term	aopo:CellTypeContext
Organ-term	aopo:OrganContext
Taxonomy	ncbitaxon:131567
Biological process	go:0008150
Biological object	pato:0001241
Biological action	pato:0000001

components, KER-specific information, and AOP assessment sections such as KE essentiality and quantitative considerations.

For the terms missing in the AOPO, we selected definitions from a wide range of other ontologies and vocabularies for the semantification of predicates and subjects, including NCI Thesaurus,³¹ NCBI of Organismal Classification, Gene Ontology, and Measurement Method Ontology. Whereas the majority of the AOP-Wiki contents are generic and can be described with well-established metadata ontologies, some properties of AOPs, KEs, and KERs could not be found, leading to the selection of more general terms, lacking detail that would be preferred. With the conversion to RDF, most of the necessary terms have been uncovered and documented, and these will be added to the AOPO.

Because the realm of AOPs includes many types of data, knowledge, repositories, and services, the development and implementation of a central, community-wide vocabulary would facilitate their integration. Since the AOPO has been developed to fill that purpose, it could be extended to include descriptions of classes and properties for all ontology terms used in the AOP-Wiki RDF. Having a central, field-wide ontology for AOPs helps maintain a high-quality vocabulary through continuous development and involvement of the community. Such an ontology would facilitate the annotation and integration of data, resources, and tools, as is done with the eNanoMapper ontology in the nanotoxicology community.⁷⁰

With the increased importance of consistent use of identifiers to integrate knowledge and data, our implementation of persistent identifiers for AOPs, KEs, KERs, stressors, chemicals, proteins, and genes will benefit the integration of the AOP-Wiki with other resources, data, and tools.⁷¹ These persistent identifiers stored in the MIRIAM registry are stable, unique, resolvable, documented, and directly linked to the corresponding entries in the databases.^{20,72}

Furthermore, our efforts have introduced additional content into the AOP-Wiki RDF through ID mappings and text-mapping for chemicals and genes, providing more ways of extrapolating the data and linking with other resources and data. While the addition of molecular descriptors of genes and proteins in KEs and KERs can provide more insights into the actual processes described, it is merely a small step toward complete machine-readable biological processes.

Although we have added molecular identifiers to increase the number of link-outs and improve the usefulness of the database, our addition of genes through textual ID mapping does introduce errors to the AOP-Wiki RDF. The automated

process on free-text content assumes good practice in writing gene symbols and names according to the HGNC guidelines.⁷³ Although HGNC strives for stable gene symbols and makes justified changes for problematic ones,⁷⁴ some gene symbols still overlap with free-text abbreviations in the AOP-Wiki and are therefore falsely recognized.

Opportunities exist to improve the AOP-Wiki machine readability by having more structured text and annotations for molecular entities, pathways, organs, species, and other biological concepts that are relevant for AOPs and not yet covered by the KE components. Text-mining tools, such as ProMiner,⁷⁵ ContentMine,⁷⁶ and PolySearch2,⁷⁷ could be implemented for extracting biological concepts and understanding associations to add more structured information to the AOP-Wiki, and facilitate the integration with other databases and tools. Once such concepts are recognized and extracted, the RDF could be extended and increase the interoperability of the AOP-Wiki with external databases, such as pathway databases.

The AOP-Wiki RDF allows for new and efficient ways of accessing the data and using it to answer questions. By loading the RDF in a SPARQL endpoint, SPARQL queries can be used to access the data and extract all necessary information. It allows complex queries across the complete AOP-Wiki database, optional filtering for any variable, and requesting an output suitable for answering the research question. Furthermore, these SPARQL queries can be executed from most coding environments as part of larger workflows or data pipelines.

It also facilitates direct linkage of databases through federated SPARQL queries, which returns information across databases with a single query. Any database with a SPARQL endpoint can be used for such questions across databases, such as WikiPathways,^{16,17} Wikidata,^{59,60} neXtProt,¹⁵ UniProt,⁵³ ChEMBL-RDF,⁷⁸ DisGeNET,^{79,80} Rhea,⁸¹ and Pathway Commons.⁸² Examples of federated queries are stored in the SPARQL examples panel in the AOP-Wiki SNORQL User Interface <https://aopwiki.rdf.bigcat-bioinformatics.org/> (Fig. 2).

Another way of using the RDF to extract AOP-Wiki content is through a web service such as the git repository linked data API constructor (grlc),⁸³ which can build a Web API on top of a SPARQL endpoint with predefined SPARQL queries. While more straightforward than SPARQL queries, the API is limited to the predefined SPARQL queries and variables implemented in these.

An advantage of creating RDF for the AOP-Wiki is the ability to link and expand AOP-Wiki content with information from other databases. For example, the Adverse Outcome Pathway Database combines knowledge from the AOP-Wiki with annotations of genes, chemicals, diseases, tissues, pathways, ontologies, and ToxCast data.^{84,85} Future work should focus on integrating such efforts by developing RDF, and thereby allow full integration of their data and tools⁸⁶ with the AOP-Wiki RDF and other databases.

In terms of compliance with Linked Data standards according to the YummyData registry, the AOP-Wiki SPARQL endpoint consistently scores above average in availability, freshness, operation, usefulness, validity, and performance.⁶⁹ With a consistent A rank, the AOP-Wiki SPARQL endpoint is placed among the 10% best scoring of the 70+ SPARQL endpoints registered in YummyData.

However, incidentally the Umaka score dropped due to slow server response, or when the service has been down for maintenance or data loading. Based on the feedback given by YummyData, a point for improvement would be supporting more response formats of the SPARQL endpoint to increase the usefulness score.

Conclusions

The implementation of compact identifiers and the development of a formal, machine-readable RDF schema make the content of the AOP-Wiki more findable and interoperable to other components of the database, and by allowing SPARQL queries and API to explore the data, the AOP-Wiki database was made accessible through new methods. Furthermore, the addition of link-outs to various chemical, gene, and protein databases, as well as the data storage in an RDF format and implementing Linked Open Data standards, has made the data more interoperable with other databases and tools.

Furthermore, the AOP-Wiki has recently introduced licenses on its content, and the code for the creation and validation of the AOP-Wiki RDF is available under the MIT license. These provide clear statements and terms of using, sharing, and modifying the content. Taken together with the addition of metadata and semantic information represented by ontology annotations, the content of the AOP-Wiki has been made more accessible and reusable. Therefore, the development of the AOP-Wiki RDF addresses all major FAIR principles.¹³

Overall, the AOP-Wiki RDF allows for new ways of exploring the data, using it in automated workflows, from coding environments, or directly through a SPARQL endpoint. With the implementation also comes the possibility to execute federated queries to combine data of multiple resources and answer more elaborate questions. For example, KEs can be linked to the results of ToxCast assays that measure the activity of the protein described, or molecular pathways can be explored for a more detailed description of mechanistic processes.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

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

Supplementary Material

Supplementary Table S1

References

1. Ankley GT, Bennett RS, Erickson RJ, et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 2010;29:730–741.
2. Krewski D, Andersen ME, Mantus E, et al. Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal* 2009;29:474–479.
3. Leist M, Ghallab A, Graepel R, et al. Adverse outcome pathways: opportunities, limitations and open questions. *Arch Toxicol* 2017;91:3477–3505.
4. Willett C. The use of adverse outcome pathways (AOPs) to support chemical safety decisions within the context of integrated approaches to testing and assessment (IATA). In: *Alternatives to Animal Testing*. H Kojima, T Seidle, H Spielmann (ed); pp. 83–90. Singapore: Springer; 2019.
5. Vinken M, Knapen D, Vergauwen L, et al. Adverse outcome pathways: a concise introduction for toxicologists. *Arch Toxicol* 2017;91:3697–3707.
6. Villeneuve DL, Crump D, Garcia-Reyero N, et al. Adverse outcome pathway (AOP) development I: strategies and principles. *Toxicol Sci* 2014;142:312–320.
7. Jeong J, Choi J. Use of adverse outcome pathways in chemical toxicity testing: potential advantages and limitations. *Environ Health Toxicol* 2017;33:e2018002.
8. Oki NO, Edwards SW. An integrative data mining approach to identifying adverse outcome pathway signatures. *Toxicology* 2016;350–352:49–61.
9. Williams AJ, Grulke CM, Edwards J, et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J Cheminform* 2017;9:61.
10. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res* 2012;40(Database issue):D136–D143.
11. Ives C, Campia I, Wang RL, et al. Creating a structured AOP knowledgebase via ontology-based annotations. *Appl In Vitro Toxicol* 2017;3:298–311.
12. Cyganiak R, Wood D, Lanthaler M. RDF 1.1 concepts and abstract syntax 2014. <https://www.w3.org/TR/rdf11-concepts/> (last accessed Aug. 12, 2020).
13. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
14. Chichester C, Karch O, Gaudet P, et al. Converting neXtProt into linked data and nanopublications. *Semantic Web* 2015; 6:147–153.
15. Zahn-Zabal M, Michel PA, Gateau A, et al. The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res* 2020;48(D1):D328–D334.
16. Waagmeester A, Kutmon M, Riutta A, et al. Using the semantic web for rapid integration of WikiPathways with other biological online data resources. *PLoS Comput Biol* 2016;12:e1004989.
17. Martens M, Ammar A, Riutta A, et al. WikiPathways: connecting communities. *Nucleic Acids Res* 2021;49(D1):D613–D621.
18. Sewell F, Gellatly N, Beaumont M, et al. The future trajectory of adverse outcome pathways: a commentary. *Arch Toxicol* 2018;92:1657–1661.
19. Wittwehr C, Aladjov H, Ankley G, et al. Adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology. *Toxicol Sci* 2017;155:326–336.
20. Juty N, Le Novère N, Laibe C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res* 2012;40(Database issue):D580–D586.
21. Martens M. marvinm2/AOPWikiRDF. Zenodo. 2021. <https://doi.org/10.5281/ZENODO.2654856>
22. Whetzel PL, Noy NF, Shah NH, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in

- software applications. *Nucleic Acids Res* 2011;39(Web Server issue);W541–W545.
23. McGuinness DL, van Harmelen F. OWL web ontology language overview. 2004. <https://www.w3.org/TR/owl-features/> (last accessed Aug. 12, 2020).
 24. DCMI Usage Board. DCMI: Dublin Core TM Metadata Element Set, Version 1.1: reference description. 2012. <https://www.dublincore.org/specifications/dublin-core/dces/> (last accessed October 3, 2020).
 25. DCMI Usage Board. DCMI: DCMI Metadata Terms. 2020. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/> (last accessed October 3, 2020).
 26. Brickley D, Guha RV. RDF Schema 1.1. 2014. <https://www.w3.org/TR/rdf-schema/> (last accessed October 3, 2020).
 27. Brickley D, Miller L. FOAF Vocabulary Specification 2014. <http://xmlns.com/foaf/spec/> (last accessed October 3, 2020).
 28. Burgoon LD. The AOPontology: a semantic artificial intelligence tool for predictive toxicology. *Appl In Vitro Toxicol* 2017;3:278–281.
 29. Mungall C, Vasilevsky N, Slater L, et al. pato-ontology/pato. 2018. <https://doi.org/10.5281/ZENODO.592861> (last accessed October 3, 2020).
 30. Hastings J, Chepelev L, Willighagen E, et al. The chemical information ontology: provenance and disambiguation for chemical data on the biological semantic web. *PLoS One* 2011;6:e25513.
 31. Sioutos N, de Coronado S, Haber MW, et al. NCI thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 2007;40:30–43.
 32. Smith JR, Park CA, Nigam R, et al. The clinical measurement, measurement method and experimental condition ontologies: expansion, improvements and new applications. *J Biomed Semantics* 2013;4:26.
 33. Miles A, Bechhofer S. SKOS simple knowledge organization system reference. 2009. <https://www.w3.org/TR/skos-reference/> (last accessed October 3, 2020).
 34. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–29.
 35. Ison J, Kalas M, Jonassen I, et al. EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics* 2013;29:1325–1332.
 36. Ciccarese P, Soiland-Reyes S, Belhajjame K, et al. PAV ontology: provenance, authoring and versioning. *J Biomed Semantics* 2013;4:37.
 37. Alexander K, Cyganiak R, Hausenblas M, et al. Describing linked datasets with the VoID vocabulary. 2011. <https://www.w3.org/TR/void/> (last accessed December 8, 2020).
 38. Albertoni R, Browning D, Cox S, et al. Data catalog vocabulary (DCAT)—Version 2. 2020. <https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/> (last accessed December 8, 2020).
 39. Bard J, Rhee SY, Ashburner M. An ontology for cell types. *Genome Biol* 2005;6:R21.
 40. Mungall CJ, Torniai C, Gkoutos GV, et al. Uberon, an integrative multi-species anatomy ontology. *Genome Biol* 2012;13:R5.
 41. Hermjakob H, Montecchi-Palazzi L, Bader G, et al. The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat Biotechnol* 2004;22:177–183.
 42. Smith CL, Eppig JT. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med* 2009;1:390–399.
 43. Baumann N. How to use the medical subject headings (MeSH). *Int J Clin Pract* 2016;70:171–174.
 44. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* 2019;47(D1):D1018–D1027.
 45. Walls RL, Guralnick R, Deck J, et al. Meeting report: advancing practical applications of biodiversity ontologies. *Stand Genomic Sci* 2014;9:17.
 46. Walls RL, Deck J, Guralnick R, et al. Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. *PLoS One* 2014;9:e89606.
 47. Gkoutos GV, Schofield PN, Hoehndorf R. The neurobehavior ontology: an ontology for annotation and integration of behavior and behavioral phenotypes. *Int Rev Neurobiol* 2012;103:69–87.
 48. Park CA, Bello SM, Smith CL, et al. The vertebrate trait ontology: a controlled vocabulary for the annotation of trait data across species. *J Biomed Semantics* 2013;4:13.
 49. Natale DA, Arighi CN, Barker WC, et al. The Protein Ontology: a structured representation of protein forms and complexes. *Nucleic Acids Res* 2011;39(Database issue):D539–D545.
 50. Degtyarenko K, de Matos P, Ennis M, et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2008;36(Database issue):D344–D350.
 51. Rosse C, Mejino JL Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform* 2003;36:478–500.
 52. Maglott D, Ostell J, Pruitt KD, et al. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2011;39(Database issue):D52–D57.
 53. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47(D1):D506–D515.
 54. Braschi B, Denny P, Gray K, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res* 2019;47(D1):D786–D792.
 55. van Iersel MP, Pico AR, Kelder T, et al. The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics* 2010;11:5.
 56. SI D. Metabolite BridgeDb ID Mapping Database (20201104). figshare. Dataset. 2020. <https://doi.org/10.6084/m9.figshare.13187585.v1> (last accessed November 5, 2020).
 57. BiGCaT. Gene/Protein BridgeDb ID Mapping Database (Ensembl 91) (Version 91) [Data set]. 2020. <https://doi.org/10.5281/zenodo.3667670> (last accessed November 5, 2020).
 58. Pence HE, Williams A. Chemspider: an online chemical information resource. *J Chem Educ* 2010;87:1123–1124.
 59. Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, et al. Wikidata as a knowledge graph for the life sciences. *Elife* 2020;9:e52614.
 60. Erxleben F, Günther M, Krötzsch M, et al. Introducing Wikidata to the Linked Data Web. In: *13th International Semantic Web Conference, October 19–23, Riva del Garda, Italy*. Cham: Springer; 2014, p. 623.
 61. Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 2012;40(Database issue):D1100–D1107.
 62. Kim S, Thiessen PA, Bolton EE, et al. PubChem Substance and Compound databases. *Nucleic Acids Res* 2016;44(D1):D1202–D1213.
 63. Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;34(Database issue):D668–D672.

64. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28;27–30.
65. Fahy E, Subramaniam S, Murphy RC, et al. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* 2009;50(Suppl);S9–S14.
66. Wishart DS, Tzur D, Knox C, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007;35(Database issue);D521–D526.
67. Hubbard T, Barker D, Birney E, et al. The Ensembl genome database project. *Nucleic Acids Res* 2002;30;38–41.
68. IDLab—Ghent University. IDLabResearch/TurtleValidator: a Turtle validator on command line and in browser. 2020. <https://github.com/IDLabResearch/TurtleValidator> (last accessed February 27, 2020).
69. Yamamoto Y, Yamaguchi A, Splendiani A. YummyData: providing high-quality open life science data. *Database (Oxford)* 2018;2018;bay022.
70. Hastings J, Jeliaskova N, Owen G, et al. eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. *J Biomed Semantics* 2015;6;10.
71. Wimalaratne SM, Juty N, Kunze J, et al. Uniform resolution of compact identifiers for biomedical data. *Sci Data* 2018;5; 180029.
72. McMurry JA, Juty N, Blomberg N, et al. Identifiers for the 21st century: how to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biol* 2017;15:e2001414.
73. Bruford EA, Braschi B, Denny P, et al. Guidelines for human gene nomenclature. *Nat Genet* 2020;52;754–758.
74. Tweedie S, Braschi B, Gray K, et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res* 2021;49(D1);D939–D946.
75. Hanisch D, Fundel K, Mevissen HT, et al. ProMiner: rule-based protein and gene entity recognition. *BMC Bioinformatics* 2005;6 Suppl 1(Suppl 1);S14.
76. Smith-Unna R, Murray-Rust P. The ContentMine scraping stack: literature-scale content mining with community-maintained collections of declarative scrapers. *D-Lib Magazine* 2014;20.
77. Liu Y, Liang Y, Wishart D. PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more. *Nucleic Acids Res* 2015;43(W1);W535–W542.
78. Willighagen EL, Waagmeester A, Spjuth O, et al. The ChEMBL database as linked open data. *J Cheminform* 2013; 5;23.
79. Piñero J, Ramírez-Angueta JM, Saüch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2020;48(D1);D845–D855.
80. Queralt-Rosinach N, Piñero J, Bravo A, et al. DisGeNET-RDF: harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases. *Bioinformatics* 2016;32;2236–2238.
81. Lombardot T, Morgat A, Axelsen KB, et al. Updates in Rhea: SPARQLing biochemical reaction data. *Nucleic Acids Res* 2019;47(D1);D596–D600.
82. Cerami EG, Gross BE, Demir E, et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 2011;39(Database issue);D685–D690.
83. Meroño-Peñuela A, Martínez C, Hoekstra R, et al. CLARIAH/grlc: January 2020 patch (Version v1.3.2). 2020. <https://doi.org/10.5281/ZENODO.3606813> (last accessed November 5, 2020).
84. Mortensen HM, Chamberlin J, Joubert B, et al. Leveraging human genetic and adverse outcome pathway (AOP) data to inform susceptibility in human health risk assessment. *Mamm Genome* 2018;29;190–204.
85. Pittman ME, Edwards SW, Ives C, et al. AOP-DB: a database resource for the exploration of Adverse Outcome Pathways through integrated association networks. *Toxicol Appl Pharmacol* 2018;343;71–83.
86. Price PS, Jarabek AM, Burgoon LD. Organizing mechanism-related information on chemical interactions using a framework based on the aggregate exposure and adverse outcome pathways. *Environ Int* 2020;138;105673.

Address correspondence to:

Marvin Martens
Department of Bioinformatics—BiGCaT
NUTRIM
Maastricht University
Universiteitssingel 50
Maastricht 6200 MD
The Netherlands

E-mail: marvin.martens@maastrichtuniversity.nl