Computational toxicology is central to the current transformation occurring in toxicology and chemical risk assessment. There is a need for more efficient use of existing data to characterize human toxicological response data for environmental chemicals in the US and Europe. The Adverse Outcome Pathway (AOP) framework helps to organize existing mechanistic information and contributes to what is currently being described as New Approach Methodologies (NAMs). AOP knowledge and data are currently submitted directly by users and stored in the AOP-Wiki (https://aopwiki.org/). Automatic and systematic parsing of AOP-Wiki data is challenging, so we have created the EPA Adverse Outcome Pathway Database. The AOP-DB, developed by the US EPA to assist in the biological and mechanistic characterization of AOP data, provides a broad, systems-level overview of the biological context of AOPs. Here we describe the recent semantic mapping efforts for the AOP-DB, and how this process facilitates the integration of AOP-DB data with other toxicologically relevant datasets through a use case example.

Keywords: semantic web, adverse outcome pathway, toxcast assays, disease, pathway, ontological mapping

Abbreviations: AOP, Adverse Outcome Pathway; AOP-DB, Adverse Outcome Pathway Database; AOP-KB, Adverse Outcome Pathway Knowledgebase; AOP-Wiki, The Collaborative Adverse Outcome Pathway Wiki; FAIR, Findable, Accessible, Interoperable, Reusable; ID, Identifier; IRI, Internationalized Resource Identifier (IRI); KE, Key Event; MIE, Molecular Initiating Event; NCBI, National Center for Biotechnology Information; NRC, National Research Council; NAMs, New Approach Methodologies; OECD, Organisation for Economic Co-operation and Development; RDF, Resource Description Framework; SPARQL, SPARQL Protocol and RDF Query Language; SQL, Standard Query Language; ToxCast, tEPA’s Toxicity Forecaster; US EPA, United States Environmental Protection Agency; W3C, World Wide Web Consortium.
INTRODUCTION

There is a need for more efficient use of existing data through improved data integration and compatibility of data structures to characterize human toxicological response data for environmental chemicals. Assessors in the US are moving towards the use of existing mechanistic data (in vitro and in silico) that provide insights into adverse outcomes in humans (National Research Council (NRC), 2007; National Research Council (NRC), 2009; National Research Council (NRC), 2010; (National Research Council (NRC), 2017), and reduced animal testing (Wheeler, 2019). The Adverse Outcome Pathway (AOP) framework helps to organize existing mechanistic information and contributes to what is currently being described as New Approach Methodologies (NAMs) (Thomas et al., 2019). The US EPA Adverse Outcome Pathway-Database (AOP-DB) is a decision support tool for risk assessors, developed by the EPA’s Center for Public Health and Environmental Assessment, which contributes to NAMs (e.g., computational toxicology tools) used for the Toxic Substances Control Act (Public Law 114–182, 2016). The AOP-DB has been made available through the Office of Science Management as a public EPA database since November 2021. Pertinent AOP-DB data is currently integrated with the CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard/chemical_lists/AOPSTRESSORS), which maps the Distributed Structure-Searchable Toxicity records to the most current list of AOP-DB stressors.

The AOP-DB integrates AOP content to help users characterize AOPs from the OECD-funded AOP-KB (https://aopkb.oecd.org/index.html) effort, where the AOP-Wiki (https://aopwiki.org/) is the primary repository for direct user submission of AOP information to the AOP-KB. Because the AOP-Wiki data is challenging to parse in its current format (Ives et al., 2017; Martens et al., 2018), the AOP-DB was developed to assist in automating and organizing AOP data, as well as integrating with publicly available datasets to allow biological and mechanistic characterization of AOPs and provide a systems-level overview of the biological context of AOPs (Mortensen et al., 2018; Pittman et al., 2018). Recent updates to AOP-DB in version 2 (Mortensen, 2021; Mortensen et al., 2021) include 280 AOPs (1,111 kEs) from the AOP-Wiki XML. The semantic mapping of AOP-DB data, described herein, extends AOP capabilities to users through the incorporation of the Research Description Framework (RDF), which creates additional ontological linkages and improves capabilities for computational analyses (Figure 1). These tools are useful to AOP users trying to retrieve information for AOP development or to understand and characterize existing AOPs. Here we describe the recent semantic mapping efforts for the AOP-DB, and how this process integrates AOP-DB data with other toxicologically relevant datasets.

As part of OpenRiskNet, a 3 years project supported by the European Commission within Horizon2020 EINFRA-22-2016 Programme, the US EPA AOP-DB was selected as an Implementation Challenge winner. The Implementation Challenge was created to select external tools for use in risk assessment to be prioritized for integration in the OpenRiskNet e-Infrastructure (https://openrisknet.org/) and foster collaborative interaction between project partners. In contribution to this effort, US EPA and Maastricht University project partners have completed the semantic mapping of several AOP-DB data tables into RDF, which is a standard model for data interchange (W3C, 2014). The application of RDF defines relationships between data objects using triplestores that include three positional statements (subject, predicate and object). The mapping of AOP-DB data to the RDF data model stores relevant AOP information in a computer-readable format, and contributes to the identification, disambiguation, and meaningful linkage of AOP data with other data structures, following FAIR (findable, accessible, interoperable, and reusable) principles (Wilkinson et al., 2019a; Wilkinson et al., 2019b).

FIGURE 1 | The OECD funded AOP-KB currently support the AOP-Wiki. The EPA AOP-DB, currently slated as a third-party tool for integration with the AOPKB 2.0, automatically and programmatically pulls AOP data from the AOP-KB XML, and extends AOP capabilities to users with semantic resources like WikiPathways and the OpenRiskNet e-infrastructure that incorporate the Research Description Framework (RDF). Integration of data across the AOP-KB (AOP-Wiki), AOP-DB, and expanding research frameworks through WikiPathways and the EU funded OpenRiskNet, creates additional ontological linkages and improves capabilities for computational analyses. These tools are useful to AOP users trying to retrieve information for AOP development, as well as those trying to understand and characterize existing AOPs.
We selected seven AOP-DB data tables for semantic integration, specifically the Gene Interaction, Biological Pathway, Toxcast Assay, Taxonomy, Chemical-Gene, Gene Info, and Key Event tables. In developing the AOP-DB RDF, we implemented the most recent version of the SQL AOP-DB (Mortensen, 2020) to map each table of interest into RDF triples. Each table was filtered using the R version 3.6 and Rstudio version 1.2.83 (R Core Team, 2020) to include only records involving a molecular initiating event (MIE) or key event (KE) that maps to a molecular identifier (e.g., gene, protein, cytokine). Code was developed to implement each record as input, modify and filter the AOP-DB table data, and output each modified record to an RDF triple. Additionally, subjects were created for Ensembl and UniProt identifiers. Ontology terms were referenced using BioPortal (Whetzel et al., 2011) in order to find the most appropriate ontology terms for each entity, in line with the AOP-Wiki RDF (Martens et al., 2021a) for optimal interoperability between the two resources. Terms were selected with the most accurate description from ontologies that are relevant to the context of the field. For the development of the AOP-DB RDF, several ontologies and consistent vocabularies have been included. Furthermore, publicly available datasets included in the AOP-DB for RDF mapping are described in detail in Mortensen et al. (2021). Table 1 provides an overview of the included ontologies and database links, including their prefix in the RDF and their corresponding Internationalized Resource Identifier (IRI).

**MATERIALS AND METHODS**

We selected seven AOP-DB data tables for semantic integration, specifically the Gene Interaction, Biological Pathway, Toxcast Assay, Taxonomy, Chemical-Gene, Gene Info, and Key Event tables. In developing the AOP-DB RDF, we implemented the most recent version of the SQL AOP-DB (Mortensen, 2020) to map each table of interest into RDF triples. Each table was filtered using the R version 3.6 and Rstudio version 1.2.83 (R Core Team, 2020) to include only records involving a molecular initiating event (MIE) or key event (KE) that maps to a molecular identifier (e.g., gene, protein, cytokine). Code was developed to implement each record as input, modify and filter the AOP-DB table data, and output each modified record to an RDF triple. Additionally, subjects were created for Ensembl and UniProt identifiers. Ontology terms were referenced using BioPortal (Whetzel et al., 2011) in order to find the most appropriate ontology terms for each entity, in line with the AOP-Wiki RDF (Martens et al., 2021a) for optimal interoperability between the two resources. Terms were selected with the most accurate description from ontologies that are relevant to the context of the field. For the development of the AOP-DB RDF, several ontologies and consistent vocabularies have been included. Furthermore, publicly available datasets included in the AOP-DB for RDF mapping are described in detail in Mortensen et al. (2021). Table 1 provides an overview of the included ontologies and database links, including their prefix in the RDF and their corresponding Internationalized Resource Identifier (IRI).

**RESULTS**

**The AOP-DB Semantic Mapping**

The AOP-DB RDF schema developed according to the methods described above resulted in the primary and secondary table
structure, as illustrated in Figure 2. The AOP-DB extends AOP-Wiki RDF with the inclusion of gene/protein, chemical, ToxCast, and biological pathway and taxonomy information. In total, the RDF contains 157 KEs, 376 NCBI genes linked to KEs, 93,449 Chemical-Gene Interactions (3,982 unique chemicals and 122 unique genes), 763,446 Protein-Protein Interactions, 1,143 ToxCast Assays, 110,833 Biological Pathways from 10 sources, and 22 taxonomies. Also, the NCBI Gene IDs were matched to 299 Ensembl IDs and 1,026 UniProt IDs. The AOP-DB RDF data tables associate the gene and protein information of AOP genes to chemical, pathway, and assay information organized within the AOP-DB (Mortensen, 2020; Mortensen, 2021).

The Key Event subjects are linked to NCBI Genes through the ‘data_1,027’ term of the EDAM ontology, which in turn is linked to pathways and assays with respectively the terms ‘pw:000001’ from the Pathway Ontology and ‘mmo:0000441’ from the Measurement Method Ontology. Furthermore, matching identifiers were linked
with ‘skos:exactMatch’, providing IRIs of Ensembl IDs, HGNC Symbols, and UniProt IDs. On the other hand, Chemical-Gene interactions, Protein-protein interactions, ToxCast assays, and Pathways have links to NCBI Gene subjects through the term ‘data_1027’ from the EDAM ontology. Finally, taxonomy is referenced by ToxCast assay and pathway subjects through the term ‘ncbitaxon:131,567’ indicating cellular organism.

The AOP-DB SPARQL Endpoint
The AOP-DB RDF can be explored through the AOP-DB SPARQL (https://aopdb.rdf.bigcat-bioinformatics.org/sparql). It allows custom SPARQL queries to return output tables in a variety of formats, where it is possible to directly combine different resources with federated SPARQL queries.

AOP-DB RDF Use Case Example
SPARQL queries can be used to query the RDF in order to answer biological and toxicological questions, such as which molecular targets (e.g. genes/proteins), chemical stressors, key events, or in vitro assays are relevant for adverse outcomes of interest. The use case examples provided herein (Supplementary 1) illustrate the utility of the AOP-DB RDF content, as well as the power of integrating these data with other diverse, external databases using federated queries. Our first use case implements the AOP-DB RDF to identify AOP-relevant molecular targets that have associated ToxCast assay targets, which has previously not been possible. The automated linkage of ToxCast assays and KEs in AOP-Wiki can serve as a prioritization tool by exploring the activation of KEs by the many chemicals that have been investigated in ToxCast. The second use case shows the integration of the AOP-DB RDF with other databases that provide access to their data through SPARQL endpoints. A single SPARQL query can be executed to extract AOP IDs, KE IDs, KE titles and protein names from the AOP-Wiki RDF, extract protein descriptions from the Protein Ontology, and the names and descriptions of pathways in WikiPathways, all based on the NCBI Gene IDs captured in the AOP-DB. Through the integration of these diverse data sources, we can effectively explore the data and build automated computational workflows to address questions of toxicological concern.

CONCLUSION
A central goal of computational toxicology is to predict and explain how the human body responds after exposure to specific xenobiotics or other chemicals in silico. This effort has been hampered by several major limiting factors, including fragmented and poorly structured data, and insufficient access to computational resources and expertise. The AOP-DB RDF and SPARQL endpoint created and discussed herein allow improved access to rigorously structured AOP data and other associated data of toxicological interest. This work improves computational organization and efficiency, through improved data integration, for toxicological and related datasets, and contributes to continued progress in computational toxicology, chemical screening and the improvement of human health risk assessment.

The AOP-DB RDF will be improved with regular data updates and continued data integration with relevant datasets. Future work includes semantic integration of AOP-DB disease-gene data, tissue-specific gene interaction networks, AOP functional single nucleotide polymorphism (SNP) and population SNP frequency information and chemical-specific datasets.

DATA AVAILABILITY STATEMENT
The original contributions presented in the study are included in the article/Supplementary Material and are also made available at: https://github.com/BiGCAT-UM/AOP-DB-RDF. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS
CTE, HM—Developed EPA AOP-DB and submitted project and data to OpenRiskNet MM—Determined project feasibility, Trained TL and facilitated project JS—modified AOP-DB RDF code TL—wrote initial RDF code for AOP-DB tables CTE Supported project, ELW—Trained TL and facilitated project TE—Supported AOP-DB RDF Research through EU funded OpenRiskNet Project.

ACKNOWLEDGMENTS
The authors gratefully acknowledge Weston Slaughter for his contribution to the queries presented in the supplemental materials, and Dr. Antony Williams for his critical comments of the earlier version of this manuscript. The training of data curators, and necessary travel were all accomplished through the identification of the EPA AOP-DB resource by Douglas Connect GmbH and with the generous contributions to this project by the EU funded OpenRiskNet.

REFERENCES
Abeyruwan, S., Vempati, U. D., Küçük-McGinty, H., Visser, U., Koleti, A., Mir, A., et al. (2014). Evolving BioAssay Ontology (BAO): Modularization, Integration and Applications. J. Biomed. Semant 5 (Suppl. 1 Proceedings of the Bio-Ontologies Spec Interest G), SS. doi:10.1186/2041-1480-5-S1-S5
Bodenreider, O. (2004). The Unified Medical Language System (UMLS): Integrating Biomedical Terminology. Nucleic Acids Res. 32 (Database issue), 267D–270D. doi:10.1093/nar/gkh061
Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., et al. (2019). Genenames.org: the HGNC and VGNC Resources in 2019. Nucleic Acids Res. 47 (D1), D786–D792. doi:10.1093/nar/gky930
Burgoon, L. (2017). Aop-Ontology. San Francisco, CA: GitHub inc.
Dumontier, M., Baker, C. J., Baran, J., Callahan, A., Chepelev, L., Cruz-Toledo, J., et al. (2014). The Semanticscience Integrated Ontology (SIO) for Biomedical Research and Knowledge Discovery. J. Biomed. Sem 5 (1), 14. doi:10.1186/2041-1480-5-14
Hastings, I., Chepelev, L., Willighagen, E., Adams, N., Steinbeck, C., and Dumontier, M. (2011). The Chemical Information Ontology: Provenance
