Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Toxicology

Ontology-based data integration for advancing toxicological knowledge

R. R. Boyles, A. E. Thessen, A. Waldrop and M. A. Haendel

Abstract

Modern toxicology is evolving to leverage data science approaches to better address complex public health concerns. Understanding the adverse health impacts of exposure is a multifactorial endeavor, which requires working across a variety of data types. These data are often siloed, and integration has required manual curation and extraction. In this review, we present the utility of adopting ontologies and semantic methods to bring disparate data into a scientifically meaningful context to drive novel scientific insights. Existing semantic standards have not been widely utilized in toxicology. Broader adoption of ontologies, together with increased data sharing, will improve a researcher's ability to integrate, navigate, and analyze vast amounts of heterogeneous data—allowing for more rapid assessment of chemical(s) safety and biological mechanisms. Recent efforts have aimed to define and realize the establishment of a data ecosystem or "commons" whereby data are shared for use by all in a common infrastructure, thereby increasing the value of government-funded data sets. Investment in making data "born interoperable" using common semantics could bring computational resources to bear on issues that are solely reliant on manual and expert assessment. Here, we introduce the concept of ontologies and present our vision for computationally enabled semantic toxicology.

Addresses

RTI International, Oregon State University, USA

Corresponding authors: Boyles, RR (rboyles@rti.org); Haendel, M (melissa@tislab.org)

Current Opinion in Toxicology 2019, **16**:67–74

This review comes from a themed issue on **Systems Toxicology**

Edited by **Stephen Edwards** and **Yu-Mei Tan**

Available online 29 May 2019

For a complete overview see the Issue and the Editorial

https://doi.org/10.1016/j.cotox.2019.05.005

1. The promise of ontological applications in toxicology

There are significant changes that have driven the use of scientific computing and data-driven approaches in modern toxicology. While data are being generated at an unprecedented rate, the importance of existing study data, such as critical animal studies, has increased the need to maximize the utility of data resources. Within biomedicine, there have been significant advances in the creation and application of data standards and ontology-driven computation that have improved data reuse and integration, as well as enabling emerging data science tools such as artificial intelligence.

An ontology is a computer readable structure that represents entities and relationships between entities [26]. Ontologies are similar to controlled vocabularies but include relationships between concepts represented using formal logic. This is a critical feature for the application of computational inference that holds such promise in maximizing existing data resources.

Ontological approaches to significantly advance information generation from diverse data have been applied in other domains. Perhaps the most high-profile application outside the biomedicine has been the Google Knowledge Graph. Released in 2012, the knowledge graph augmented the traditional text-based search strategy to provide more contextualized information to users, such as returning the hours and location of a restaurant instead of just a link to the web page. A more culturally relevant example might be IBM's Watson, which used knowledge graphs and machine learning to interpret and answer questions on the *Jeopardy!* game show. In 2018, Google released data search, which leverages the ontology, schema.org, to allow users to search hundreds of data repositories in one query. Ontologies have enabled smart home Internet of Things services [59] and mining of electronic health record (EHR) records to identifying single nucleotide polymorphism (SNPs) associated with type 2 diabetes mellitus [51]. All of these examples illustrate the potential for ontologies to make semantic interpretation possible at the largest scales and with almost any data type.

In this article, we offer a descriptive and prescriptive vision for how semantic technologies can address this paucity in the toxicological sciences. Ultimately, we argue that the large-scale adoption of ontologies across the field will help translate unprecedented volumes of data into more informed scientific, public health, and regulatory decisions.

2. Toxicology as a big data discipline

Much has been written about the explosion of biomedical data that has been largely driven by the genomic revolution [24]. In addition to the increasing availability and volume of genomic data, toxicologists have seen a dramatic increase in data through the adoption of high-throughput assays. The importance of these assays for increasing the speed, decreasing the cost, and reducing the burden of animal testing for chemical safety assessment was notably described in the NAS Toxicity Testing in the 21st Century: A Vision and a Strategy [48]. Within toxicology, this has meant a rapid increase in the heterogeneity as well as the volume of data. Modern toxicology now includes data from high-throughput genomic, proteomic, metabolomic, pharmacokinetic, behavioral, dietary, microbiome, expression assays, and wearables to prioritize chemical testing and to garner insight into environmental contributors to the etiology of disease [1,5,7,20,25,63,67].

Understanding how genetics and the environment synergistically combine and contribute to overall health is necessarily cross disciplinary. Such assessment will rely on our ability to perform data integration across sources and data modalities to access lines of evidence for precursor events and pathways—which are currently largely performed by expert, manual assessment. While expert scientific judgment will always play a critical role in advancing our understanding of toxicological data [17], continued reliance on manual human interpretation limits the possibilities for data-driven discovery. New approaches that leverage ontology-driven computation alongside expert interpretation have the potential to drive new innovations [8].

In the risk assessment domain, regulatory bodies, such as the US Food and Drug Administration [10] and the EU European Chemicals Agency [9], rely on structured electronic data to support data-intensive assessments of chemical and pharmaceutical safety [18,22]. These regulatory requirements to assess a growing number of chemicals and drugs, coupled with animal-testing reduction goals, have increased the need to derive insights from existing *in vivo*, *in vitro*, and high-throughput assay data [28,56,60]. Most recently, the Toxic Substances Control Act regulatory reform under the Frank R. Lautenberg Chemical Safety for the 21st Century Act and its administration by the US Environmental Protection Agency has placed an increased emphasis on gathering and assessment of data in hazard determinations [37].

In short, there is a current tension that is somewhat unique to toxicology; on the one hand, we are generating more data and more data types than ever before that are relevant to understanding human health effects, and on the other hand, the regulatory bodies, new exposure types/scenarios, and ethical requirements are needing us to derive much more knowledge from limited resources. There is therefore a significant and broad opportunity cost in continuing to underutilize existing data assets. Accessing and enabling data integration could support novel investigations into health impacts of chemical mixtures [29], causal inference [21], weight of evidence [40], systems toxicology approaches [34,38], predictive modeling for chemical alternatives [6], and exposome analytics [4,54] to name a few. At present, these applications and the larger ability of environmental health scientists to capitalize on data science approaches such as machine learning/artificial intelligence are hampered by the paucity of frameworks for integrating and interpreting across siloed toxicological data sources.

3. Deriving value from data

A siloed data set, isolated from related information and perhaps lacking comprehensive descriptors, has limited value [69]. In risk assessment, a data resource, in addition to the statistical analyses and metadata needs that all scientific data share, requires a "weight of evidence" approach that accumulates support from multiple independent sources before a fact is accepted [2,11,65]. Corroborating findings across studies or data sets can increase the value of a data set by contributing to the weight of evidence. For example, a toxicology study identifying the interference of a substance with a biological process becomes more useful when combined with a study showing the genes that control that process, a study showing the phenotypes that result from that process, and another study linking those phenotypes to a disease [11,55]. Data become increasingly more valuable to determine the weight of the evidence when accompanied by sufficient descriptors [23,30]; it is therefore especially critical to utilize community standards and well-formed identifiers to identify research data elements [43]. In this context, ontologies such as the Scientific Evidence and Provenance Ontology [12] can be better utilized to determine the degree of corroboration. It is this contextualization of data that enables meaningful integration and derivation knowledge, making the data "valuable" in a way that advances public and environmental health.

Knowledge graphs are an emerging form of database structure that can accommodate large, complicated data sets for advanced search and support artificial intelligence applications. Google adopted this approach in 2012, and it has long been the data infrastructure behind Wikipedia [52]. One such instance of a biomedical knowledge graph is SciGraph, and it can be accessed via the Monarch Initiative Web site and application programming interface (API) [46]. Currently, most toxicology information is not a part of this knowledge graph, with the exception of a handful of resources including the Comparative Toxicogenomics Database [16,47].

In order for the toxicology community to leverage the full potential of their data by joining this biomedical data ecosystem, further format, nomenclature, metadata, and identifier standardization need to be adopted. Adoption of these standards will allow transformation of the data into common knowledge graphs. One can imagine a future ecosystem of data resources that supports exchange and integration through common semantics, making the knowledge graph worth more than the sum of its parts. In addition, integrating toxicology data into a broader data ecosystem could spur innovation in tooling for the field by removing the challenge of meaningful data availability for analysts and software developers.

4. Data standards for comparable and consistent classification

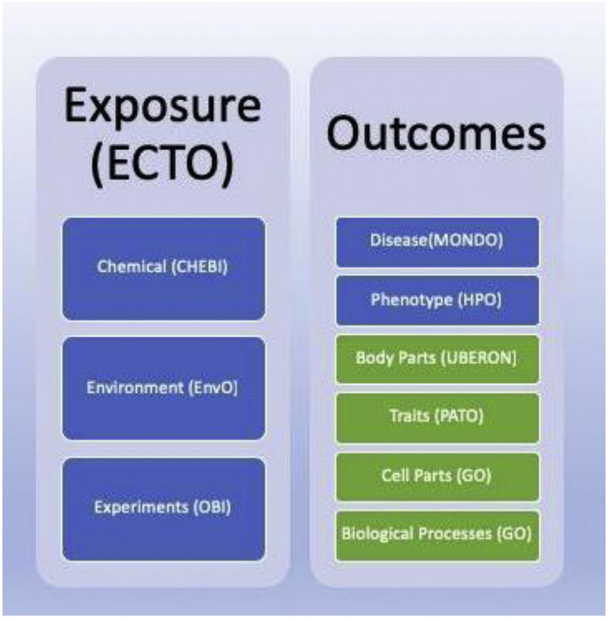
The development of useful, sustainable, and community-driven data standards is a major undertaking that requires careful thought. Standards development is not thought of by many as "science" yet forms much of the bedrock of scientific data resources [23,30]. The standards process can be driven by a dedicated few, but development is a community effort. It takes equal parts domain knowledge, discipline, social engineering, and long-term sustainability planning. Standards come in three different flavors:

1. Format: This refers to the actual structure of the data files and what format they are in. Should data be in CSV or TXT files? Specialty disciplines need specialty formats, such as for genomic data (VCF) or imaging cytometry (.ics). For example, for a CSV file format, how should the columns and rows be structured? Is there a header row? How will dates be reported? Will there be a readMe or a meta file and is this file computationally actionable given the data file?
2. Content: This refers to the file contents. Some standards mandate a specific suite of metadata to accompany a specific type of data. For example, a data set containing patient's cholesterol measurements could also require the patient's age, fasting status, and medication status. Content is often numeric, and ranges are often defined for specific variables.
3. Identifiers/vocabularies: This refers to the exact terms used to describe what is being represented in the data file. For example, a standard may require that a variable representing the concentration of a chemical in water represent that chemical using a Uniform Resource Identifier with the units given in a readMe. This is the type of standard that adds semantics to the data if the identifiers used are part of an ontology (Uniform Resource Identifier) or knowledge graph.

Figure 1

Integration of disparate data. Data increase in scientific value as they are combined and compared across types and sources. This integration of data is often performed manually, can be custom scripted, or through some combination of manual work and scripting. A few large data resources exist for the community, but there is currently a lack of a generalizable data ecosystem that would further increase data utility.

Figure 2



Building an exposure and outcomes representation using ontologies. A simple approach to using existing ontologies to represent environmental exposure. Exposures can be built using terms from the CHEBI, ontology for biomedical investigations (OBI), and EnvO ontologies (blue). Outcomes can be built using MONDO and HPO (blue). Phenotypes are linked to diseases, but they are also composed of other ontologies that describe anatomy (UBERON), traits phenotype and trait ontology (PATO), cell parts (GO) and biological processes (GO). In this way, knowledge representation in toxicology can leverage existing work. CHEBI, Chemical Entities of Biological Interest.

While the need for data quality, data descriptions (i.e. metadata or data dictionaries), and file formats is well-acknowledged, limitations to deriving value from a data set, the application of clear and defined labels and identifiers, or their use in common data models across sources, within data sets is less widely emphasized. Ontologies provide a mechanism for standard descriptors of data. Many of the standards and identifiers that are needed in toxicology already exist within the larger bio-ontology community, having been developed by other related user communities. The Gene Ontology (GO) [62] describes genes, proteins, biological processes, and cell parts. Uberon describes anatomical parts, the Cell Ontology cell types. The Chemical Entities of Biological Interest [19,42] describes chemicals. The Human Phenotype Ontology (HPO) [35,36] describes phenotypes. The Monarch Disease Resource (MONDO) [45] describes diseases and integrates them from different sources; similarly, the Environments, Exposures, and Treatments Ontology describes and integrates exposure terms. The aforementioned are only a few of the ontologies relevant for toxicology. The usefulness of these ontologies relies not only on the creation of persistent, unique identifiers for concepts but also in the relationships of the concepts to each other. For example, in MONDO, "heart lymphoma" is a subclass of "heart disease" with seven synonyms. This means that a user can search for "heart disease" and be reasonably assured to get all the data relating to any synonym or a subclass thereof. Concepts can also be related across ontologies. For example, MONDO and HPO classes are related to each other using the Relations Ontology to link phenotypes to diseases, such as linking heart murmur and long limbs to Marfan Syndrome. This ontological structure provisioned with empirically derived assertions is called a knowledge graph [3].

One of the most important things in managing data for semantic integration is good identifier hygiene. In Toxicology, chemicals are identified using their CASRN, and genes can be organized by their NCBI Gene IDs. Diseases are often represented using MeSH and OMIM identifiers and are harmonized in MONDO. However, it is not only the identifiers of these data types but also the relationships that are used to link them. For example, there are many relationships between genes and diseases, conflating causality with modifier status can lead to very incorrect conclusions. Use of entity identifiers and alignment of the relationships between them means that these data can be interlinked between the many databases within the Toxicology Data Network [49]. If the identifiers and their relationships both follow the same standard (such as use of the biolink-model standard), then these identifiers provide the ability to create a knowledge graph. The more data that comply with these standards, the more data can be brought into the knowledge graph and the more valuable the entire system becomes.

Exploiting the promise of big data in toxicology relies on our ability to align data across subdisciplines and sources. For example, over 20 years ago, veterinarians noticed a dramatic increase in feline hyperthyroidism in house cats [44,50,53]. This has since been linked to household exposure to brominated flame retardants and provided mechanistic support to the growing investigation of effects on thyroid hormone levels in people, with particular concern on understanding effects of endocrine disruption in young children, whose exposure to these flame retardants through dust is higher than older children and adults [41]. The study of these effects is still an active area of research decades after the initial veterinary diagnosis was made and relies on connections that were made by experts. Ontology-based data integration is one tool that could speed up these linkages across domains and bring potential public health issues to light (see Figure. 1).

5. Ontologies for data integration

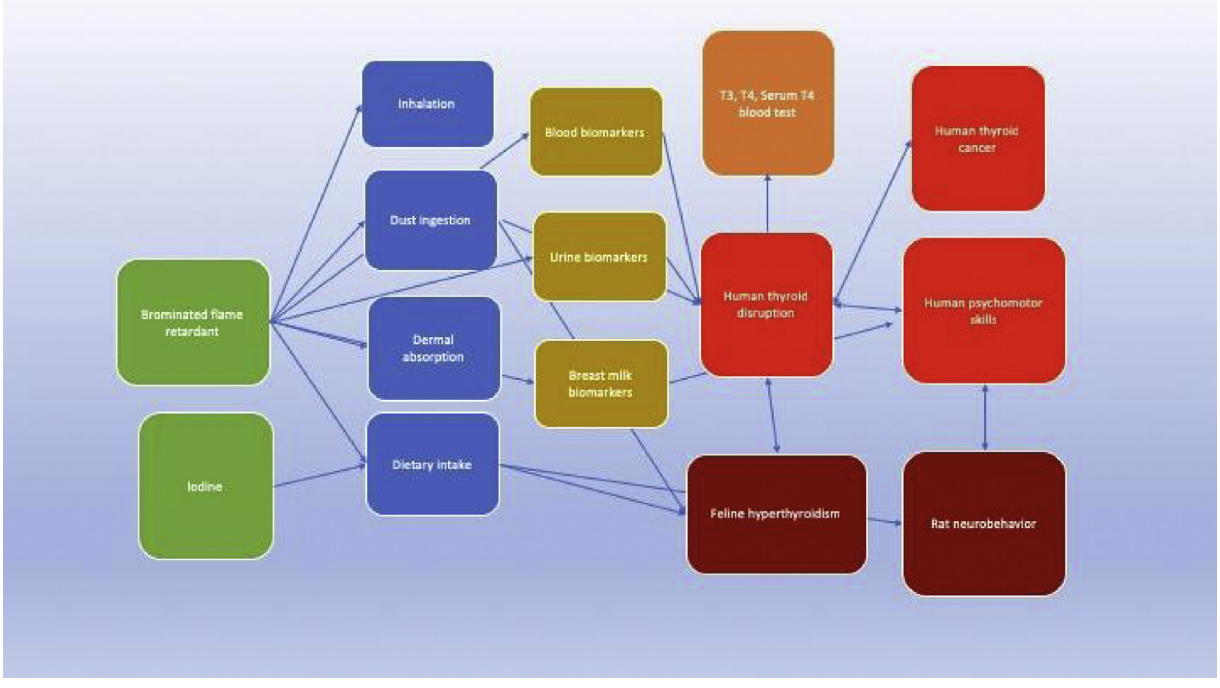
Large-scale toxicology data integration efforts are still relatively young [15]. A first critical step and a necessary

precondition toward leveraging the power of machine learning/artificial intelligence to predict chemical toxicity is the development of toxicology-focused ontologies which will allow the integration of data across years of studies and high-throughput tox assays (Wambaugh et al., 2013) [64]. A central challenge to the development of such ontologies continues to be the heterogeneity of data types in the field, which includes concepts such as chemical agent, biological pathways, genetic interactions, dosages, duration, and treatment conditions. Additionally, like data, ontologies can be of variable quality. The Open Biological and Biomedical Ontologies Foundry is one group that tries to establish quality standards for these ontologies and critically emphasizes standards that allow ontologies to be integrated in a building block approach (Figure 2) [57].

Much like ontology-based tools in other disciplines (e.g. Monarch), toxicology ontologies such as OpenTox [61], eTOX [13], eNanoMapper [32], and the adverse outcome pathway (AOP) ontology [31] combine existing ontologies such as GO, the Disease Ontology, and phenotype ontologies (e.g. HPO) within a larger framework linking chemical agents and/or cellular events to downstream biological events. For example, the OpenTox ontology, which was among the first centralized efforts to create a toxicology-specific ontology, connects chemical agent endpoints to genic, cellular, and organismal outcomes. More recently developed ontologies have focused more specifically on facilitating computational prediction. For example, the recently developed AOP ontology formalizes the structure of Adverse Outcome Pathways (Wittwehr et al., 2017) [68] with the aim of centralizing the collection of AOPs across toxicity studies. The AOP ontology was specifically designed to integrate data across high-throughput tox screens and to facilitate the computational prediction of new AOPs as new data are added.

A central goal for both ontologies is to marry the increasing capacity for toxicology data generation to artificial intelligence methods made possible by increasing computational capacity. The OpenTox, eTOX, and AOP projects have since created centralized data repositories with the OpenToxipedia [61], eTOXsys [58], and the AOPKnowledgebase [31], respectively. Data integration tools such as Monarch, rely on large semantically integrated data sets to provide cross-species queries of genotypes and phenotypes. Despite the utility of these existing data resources, the majority of toxicological data remains outside of a database or repository and lacks semantic standardization. Data sharing efforts coupled with semantic standards have the potential to significantly increase the knowledge that can be generated from existing resources. It is not difficult to imagine a nearly instantaneous

Figure 3



Mapping toxicology knowledge to the biomedical knowledge graph. Environmental exposures (green) can be linked to their route of exposure (blue), any biomarkers (mustard), and empirically derived (from study results) outcomes (red and orange). This graph can expand to represent locations, occupations, genes, and functions, which adds value to the original toxicological conclusion.

computational integration of *in vitro* bioassay data with results from observational studies [27] and specific local population and exposure data. This includes use of *in vitro* bioassay data in combination with exposure estimates to derive a quantitative assessment of risk [33].

In the future, we imagine that further adoption of these approaches will lead to support for the development of queries and tools that are capable of crossing systems and domains. Fully semantic toxicology data could join the larger biomedical knowledge graph via links from exposures to the outcome phenotype or disease (Figure 3).

6. Conclusion

The field of toxicology has found itself immersed in problems that go hand-in-hand with taking the leap into big data. Other disciplines have successfully used ontologies to help mitigate these problems, but this can only happen after a community-driven effort to develop and maintain data standards. Currently available semantic resources for toxicology show lack of community involvement and little uptake. Much of the semantic technology needed by toxicology has already been developed, but bringing the research benefits to full fruition requires a few, highly-collaborative components that only the user community can provide. Once in place, barriers, such as aggregating toxicological endpoints across organisms, will be much easier overcome. The cost of developing community standards is dwarfed by the research benefits of joining the larger biomedical knowledge graph and achieving toxicology's ultimate goal which is to address current and emerging public health concerns.

Acknowledgements

This work was supported by the National Institutes of Health [NIH NHLBI, 1 OT3 HL147154-01; NIH OD, 1 OT3 OD025464-01 S^2; NIH OD, Grant #5R24OD011883; NIEHS, ILS180706].

Conflict of interest statement

Nothing declared.

References

1. Adeleye Yeyejide, Andersen Melvin, Clewell Rebecca, Davies Michael, Dent Matthew, Edwards Sue, Fowler Paul, *et al.*: **Implementing toxicity testing in the 21st Century (TT21C): making safety decisions using toxicity pathways, and progress in a prototype risk assessment**. *Toxicology* 2015, **332**(June):102–111.

2. Ågerstrand Marlene, Anna Beronius: **Weight of evidence evaluation and systematic review in EU chemical risk assessment: foundation is laid but guidance is needed**. *Environ Int* 2016, **92–93**(July):590–596.

3. Bakal Gokhan, Talari Preetham, Kakani Elijah V, Kavuluru Ramakanth: **Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative Relations**. *J Biomed Inform* 2018, **82**(June):189–199.

4. Barouki Robert, Audouze Karine, Coumoul Xavier, Demenais Florence, Gauguier Dominique: **Integration of the human exposome with the human genome to advance medicine**. *Biochimie* 2018, **152**(September):155–158.

5. Bell Shannon M, Xiaoqing Chang, Wambaugh John F, Allen David G, Mike Bartels, Brouwer Kim LR, Casey Warren M, *et al.*: **In vitro to in vivo extrapolation for high throughput prioritization and decision making**. *Toxicol Vitro Int J Pub Assoc BIBRA* 2018, **47**(March):213–227.

6. Bell Shannon M, Phillips Jason, Alexander Sedykh, Tandon Arpit, Sprankle Catherine, Morefield Stephen Q, Shapiro Andy, *et al.*: **An integrated chemical environment to support 21st-century toxicology**. *Environ Health Perspect* 2017, **125**. 054501.

7. Bergmann Alan J, North Paula E, Vasquez Luis, Bello Hernan, Gastañaga Ruiz Maria Del Carmen, Anderson Kim A: **Multi-class chemical exposure in rural Peru using silicone wristbands**. *J Expo Sci Environ Epidemiol* 2017, **27**:560–568.

8. Biomedical Data Translator Consortium: **Toward a universal biomedical data translator**. *Clin Transl Sci* 2019, **12**:86–90.

9. Boer, Jacob de, van Bavel Bert: **"European 'REACH' (registration, evaluation, authorisation and restriction of chemicals) program**. *J Chromatogr, A* 2009, **1216**:301.

10. Briggs Katharine: **Making sense of SEND; the standard for exchange of nonclinical data**. *Regul Toxicol Pharmacol: RTP (Regul Toxicol Pharmacol)* 2017, **91**:77–85.

11. Browne Patience, Noyes Pamela D, Casey Warren M, Dix David J: **"Application of adverse outcome pathways to U.S. EPA's endocrine disruptor screening program**. *Environ Health Perspect* 2017, **125**. 096001.

12. Brush Matthew H, Kent Shefchek, Haendel Melissa: **SEPIO: a semantic model for the integration and analysis of scientific evidence**. *CEUR Workshop Proceed* 2016, **1747**. https://ohsu.pure.elsevier.com/en/publications/sepio-a-semantic-model-for-the-integration-and-analysis-of-scient.

13. Cases M, Pastor M, Sanz F: **The eTOX library of public resources for in silico toxicity prediction**. *Mol Inform* 2013, **32**:24–35.

15. Cimino MC, Auletta AE: **Availability of the GENE-TOX database on the national library of medicine TOXNET system**. *Environ Mol Mutagen* 1993, **21**:306–307.

16. Davis Allan Peter, Grondin Cynthia J, Johnson Robin J, Sciaky Daniela, King Benjamin L, Roy McMorran, Wiegers Jolene, Wiegers Thomas C, Mattingly Carolyn J: **The comparative Toxicogenomics database: update 2017**. *Nucleic Acids Res* 2017, **45**:D972–D978.

17. Davis Allan Peter, Wiegers Thomas C, Rosenstein Michael C, Murphy Cynthia G, Mattingly Carolyn J: **The curation paradigm and application tool used for manual curation of the scientific literature at the comparative Toxicogenomics database**. *Database J Biol Databases Curation* 2011, **2011**(September):bar034.

18. Declerck Gunnar, Bousquet Cédric, Jaulent Marie-Christine: **Automatic generation of MedDRA terms groupings using an ontology**. *Stud Health Technol Inf* 2012, **180**:73–77.

19. Degtyarenko Kirill, de Matos Paula, Ennis Marcus, Hastings Janna, Martin Zbinden, McNaught Alan, Alcántara Rafael, Darsow Michael, Guedj Mickaël, Ashburner Michael: **ChEBI: a database and ontology for chemical entities of biological interest**. *Nucleic Acids Res* 2008, **36**(Database issue):D344–D350.

20. Du Yanjun, Wang Qin, Sun Qinghua, Zhang Ting, Li Tiantian, Yan Beizhan: **Assessment of PM2.5 monitoring using Micro-PEM: a validation study in a city with elevated PM2.5 levels**. *Ecotoxicol Environ Saf* 2019, **171**(April):518–522.

21. Fedak Kristen M, Bernal Autumn, Capshaw Zachary A, Gross Sherilyn: **Applying the bradford hill criteria in the 21st Century: how data integration has changed causal inference in molecular epidemiology**. *Emerg Themes Epidemiol* 2015, **12**(September):14.

22. Foth H, Hayes Aw: **Concept of REACH and impact on evaluation of chemicals**. *Hum Exp Toxicol* 2008, **27**:5–21.

23. Greenberg Jane: **Big metadata, smart metadata, and metadata capital: toward greater synergy between data science and metadata**. *J Data Inform Sci* 2017, **2**:19–36.

24. Green Eric D, Watson James D, Collins Francis S: **Human genome project: twenty-five years of big biology**. *Nature* 2015, **526**:29–31.

25. Griffin JL, Bollard ME: **Metabonomics: its potential as a tool in toxicology for safety assessment and data integration**. *Curr Drug Metabol* 2004, **5**:389–398.

26. Gruber Thomas R: **Toward principles for the design of ontologies used for knowledge sharing?** *Int J Hum Comput Stud* 1995, **43**:907–928.

27. Gwinn Maureen R, Axelrad Daniel A, Bahadori Tina, Bussard David, Cascio Wayne E, Deener Kacee, Dix David, Thomas Russell S, Kavlock Robert J, Burke Thomas A: **Chemical risk assessment: traditional vs public health perspectives**. *Am J Public Health* 2017, **107**:1032–1039.

28. Hamm Jon, Sullivan Kristie, Clippinger Amy J, Strickland Judy, Bell Shannon, Bhhatarai Barun, Blaauboer Bas, *et al.*: **Alternative approaches for identifying acute systemic toxicity: moving from research to regulatory testing**. *Toxicol Vitro Int J Publ Assoc BIBRA* 2017, **41**(June):245–259.

29. Hernández Antonio F, Tsatsakis Aristidis M: **Human exposure to chemical mixtures: challenges for the integration of toxicology with epidemiology data in risk assessment**. *Food Chem Toxicol: An Int J Publ Br Indust Biol Res Assoc* 2017, **103**(May):188–193.

30. International Society for Biocuration: **Biocuration: distilling data into knowledge**. *PLoS Biol* 2018, **16**. e2002846.

31. Ives Cataia, Campia Ivana, Wang Rong-Lin, Wittwehr Clemens, Edwards Stephen: **Creating a structured AOP knowledgebase via ontology-based annotations**. *Appl Vitro Toxicol* 2017, **3**: 298–311.

32. Jeliazkova Nina, Chomenidis Charalampos, Doganis Philip, Fadeel Bengt, Grafström Roland, Hardy Barry, Hastings Janna, *et al.*: **The eNanoMapper database for nanomaterial safety information**. *Beilstein J Nanotechnol* 2015, **6**(July):1609–1634.

33. Kavlock Robert J, Bahadori Tina, Barton-Maclaren Tara S, Gwinn Maureen R, Rasenberg Mike, Thomas Russell S: **Accelerating the pace of chemical risk assessment**. *Chem Res Toxicol* 2018, **31**:287–290.

34. Knudsen Thomas B, Keller Douglas A, Sander Miriam, Carney Edward W, Doerrer Nancy G, Eaton David L, Fitzpatrick Suzanne Compton, *et al.*: **FutureTox II: in vitro data and in silico models for predictive toxicology**. *Toxicol Sci Off J Soc Toxicol* 2015, **143**:256–267.

35. Köhler Sebastian, Carmody Leigh, Vasilevsky Nicole, Julius O, Jacobsen B, Danis Daniel, Gourdine Jean-Philippe, Gargano Michael, *et al.*: **Expansion of the human phenotype ontology (HPO) knowledge base and resources**. *Nucleic Acids Res* 2019, **47**:D1018–D1027.

36. Köhler Sebastian, Vasilevsky Nicole A, Engelstad Mark, Foster Erin, McMurry Julie, Aymé Ségolène, Baynam Gareth, *et al.*: **The human phenotype ontology in 2017**. *Nucleic Acids Res* 2017, **45**:D865–D876.

37. Krimsky Sheldon: **The unsteady state and inertia of chemical regulation under the US toxic substances control Act**. *PLoS Biol* 2017, **15**. e2002404.

38. Leung Maxwell CK, Jimmy Phuong, Baker Nancy C, Sipes Nisha S, Klinefelter Gary R, Martin Matthew T, McLaurin Keith W, *et al.*: **Systems toxicology of male reproductive development: profiling 774 chemicals for molecular targets and adverse outcomes**. *Environ Health Perspect* 2016, **124**:1050–1061.

40. Linkov Igor, Massey Olivia, Keisler Jeff, Rusyn Ivan, Hartung Thomas: **"From 'weight of evidence' to quantitative data integration using multicriteria decision analysis and bayesian methods**. *ALTEX - Alter Anim Experiment* 2015, **32**: 3–8.

41. Malliari Eleftheria, Olga-Ioanna Kalantzi: **"Children's exposure to brominated flame retardants in indoor environments - a review**. *Environ Int* 2017, **108**(November):146–169.

42. Matos Paula de, Alcántara Rafael, Dekker Adriano, Ennis Marcus, Hastings Janna, Haug Kenneth, Inmaculada Spiteri, Turner Steve, Steinbeck Christoph: **Chemical entities of biological interest: an update**. *Nucleic Acids Res* 2010, **38**(Database issue):D249–D254.

43. McMurry Julie A, Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte, Mélanie Courtot, *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.

44. Mensching Donna A, Margaret Slater, Scott John W, Ferguson Duncan C, Beasley Val R: **The feline thyroid gland: a model for endocrine disruption by polybrominated diphenyl ethers (PBDEs)?** *J Toxicol Environ Health Part A* 2012, **75**(4): 201–212.

45. MONDO: ***MONDO: Monarch disease ontology***. https://www.ebi.ac.uk/ols/ontologies/mondo; 2019.

46. Mungall Christopher J, McMurry Julie A, Köhler Sebastian, Balhoff James P, Borromeo Charles, Brush Matthew, Carbon Seth, *et al.*: **The Monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species**. *Nucleic Acids Res* 2017, **45**:D712–D722.

47. Muñoz Emir, Novácek Vít, Vandenbussche Pierre-Yves: **Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models**. *Briefings Bioinf* 2017, **August**. https://doi.org/10.1093/bib/bbx099.

48. National Research Council: *Toxicity testing in the 21st Century: a vision and a strategy*. Washington, DC: The National Academies Press; 2007.

49. NLM: *Toxicology, environmental health, chemical information | NLM | NIH | HHS." national library of medicine*. 2018. https://envirotoxinfo.nlm.nih.gov/index.html. Accessed 23 May 2018.

50. Norrgran Engdahl J, Bignert A, Jones B, Athanassiadis I, Bergman Å, Weiss JM: **"Cats' internal exposure to selected brominated flame retardants and organochlorines correlated to house dust and cat Food**. *Environ Sci Technol* 2017, **51**: 3012–3020.

51. Pathak Jyotishman, Kiefer Richard C, Bielinski Suzette J, Chute Christopher G: **"Mining the human phenome using semantic web technologies: a case study for type 2 Diabetes." AMIA ... Annual symposium proceedings/AMIA symposium**. *AMIA Symposium* 2012, **2012**(November):699–708.

52. Paulheim Heiko: **"Knowledge graph refinement: a survey of approaches and evaluation methods." edited by philipp cimiano**. *Semantic Web* 2016, **8**:489–508.

53. Peterson Mark E: **Animal models of disease: feline hyperthyroidism: an animal model for toxic nodular goiter**. *J Endocrinol* 2014, **223**:T97–T114.

54. Rappaport Stephen M, Smith Martyn T: **Epidemiology. Environment and disease risks**. *Science* 2010, **330**:460–461.

55. Rooney Andrew A, Boyles Abee L, Wolfe Mary S, Bucher John R, Thayer Kristina A: **Systematic review and evidence integration for literature-based environmental health science assessments**. *Environ Health Perspect* 2014, **122**:711–718.

56. Schönfelder Gilbert, Grune Barbara, Hensel Andreas: **"Animal welfare: europe's first '3Rs' governmental centre**. *Nature* 2015, **527**:38.

57. Smith Barry, Ashburner Michael, Rosse Cornelius, Bard Jonathan, Bug William, Werner Ceusters, Louis J, Goldberg, *et al.*: **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration**. *Nat Biotechnol* 2007, **25**:1251–1255.

58. Steger-Hartmann Thomas, Pognan Francois: **Improving the safety assessment of chemicals and drug candidates by the**

integration of bioinformatics and chemoinformatics data. *Basic Clin Pharmacol Toxicol* 2018, **123**(Suppl 5):29–36.

59. Tao Ming, Ota Kaoru, Dong Mianxiong: **Ontology-based data semantic management and application in IoT- and cloud-enabled smart homes**. *Future Gener Comput Syst: FGCS* 2017, **76**:528–539.

60. Taylor Katy, Stengel Wolfgang, Casalegno Carlotta, Andrew David: **Experiences of the REACH testing proposals system to reduce animal testing**. *ALTEX* 2014, **31**:107–128.

61. Tcheremenskaia Olga, Benigni Romualdo, Nikolova Ivelina, Jeliazkova Nina, Escher Sylvia E, Batke Monika, Baier Thomas, *et al.*: **OpenTox predictive toxicology framework: toxicological ontology and semantic media wiki-based OpenToxipedia**. *J Biomed Semant* 2012, **3**(Suppl 1 April):S7.

62. The Gene Ontology Consortium: **The gene ontology resource: 20 Years and still GOing strong**. *Nucleic Acids Res* 2018, **47**. https://doi.org/10.1093/nar/gky1055.

63. Truong Lisa, Reif David M, St Mary Lindsey, Geier Mitra C, Truong Hao D, Tanguay Robert L: **Multidimensional in vivo hazard assessment using zebrafish**. *Toxicol Sci Off J Soc Toxicol* 2014, **137**:212–233.

64. Wambaugh John F, Woodrow Setzer R, Reif David M, Gangwal Sumit, Mitchell-Blackwood Jade, Arnot Jon A, Joliet Olivier, *et al.*: **High-throughput models for exposure-based chemical prioritization in the ExpoCast project**. *Environ Sci Technol* 2013, **47**:8479–8488.

65. Weed, Douglas L: **Weight of evidence: a review of concept and methods**. *Risk Anal Off Publ Soc Risk Anal* 2005, **25**:1545–1557.

67. Wilmes Anja, Limonciel Alice, Aschauer Lydia, Moenks Konrad, Bielow Chris, Leonard Martin O, Hamon Jeremy, *et al.*: **Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced cell stress**. *J Proteomics* 2013, **79**(February):180–194.

68. Wittwehr Clemens, Aladjov Hristo, Ankley Gerald, Byrne Hugh J, de Knecht Joop, Heinzle Elmar, Klambauer Günter, *et al.*: **How adverse outcome pathways can aid the development and use of computational prediction models for regulatory toxicology**. *Toxicol Sci Off J Soc Toxicol* 2017, **155**:326–336.

69. Kim Gang-Hoon, *et al.*: **Big-data applications in the government sector**. *Commun ACM* 2014, **57**:78–85. https://doi.org/10.1145/2500873.