Semantic Web 0 (0) 1
IOS Press

James P. McCusker [a,1], John S. Erickson [a] and
Katherine Chastain [a] and Sabbir Rashid [a] and
Rukmal Weerawarana [a] and Marcello Bax [a] and
Deborah L. McGuinness [a]

[a] *Computer Science, Rensselaer Polytechnic Institute, Troy, NY, US*
E-mails: mccusj@cs.rpi.edu, erickj4@rpi.edu, chastk@rpi.edu, rashis2@rpi.edu, weerar@rpi.edu, baxm2@rpi.edu, dlm@cs.rpi.edu

What is a Knowledge Graph?

Abstract. Knowledge graphs have enjoyed a resurgence in research interests after the development of several commercial projects, such as Google's knowledge graph. However, the use of the term has evolved and now may refer to a wide range of graphs, that may not include clear and unambiguous definitions or references. To better provide clarity to knowledge graph research, we survey the literature for current efforts that may inform a knowledge graph definition, and then use that review along with our work to synthesize a definition that is relevant and informative to current knowledge graph research, while constraining the research space that may be considered a knowledge graph. We define a knowledge graph as "A graph, composed of a set of assertions (edges labeled with relations) that are expressed between entities (vertices), where the meaning of the graph is encoded in its structure, the relations and entities are unambiguously identified, a limited set of relations are used to label the edges, and the graph encodes the provenance, especially justification and attribution, of the assertions." We evaluate a wide variety of knowledge resources, graphs, and ontologies to determine if they qualify under our definition, and find that while expressing knowledge as a graph structure and unambiguous denotation of entities and relations in the graph are common, it is less common to trace provenance of encoded knowledge, and less common to constrain the relations used when expressing that knowledge. We created our Knowledge Graph Catalog to support this effort, and make it available to the public to search and contribute new knowledge graphs.

Keywords: Knowledge Graphs

1. Introduction

Google introduced its Knowledge Graph project in 2012 [1] in order to enhance their search result quality, but it has also reignited interest in knowledge graph research. They have leveraged existing knowledge graphs, such as DBpedia and Freebase, and also have opened up the process of contributing to the graph by ingesting linked data, RDFa, and microdata formats from the Web pages they index, based on the vocabularies published by schema.org. The success of the Google Knowledge Graph, and its use of semantic technologies, has led to a resurgence in the use of the term in semantic research to describe similar projects. However, the term "knowledge graph" remains underspecified, and in many cases, simply refers to any directed labeled graph. The pre-Semantic Web conceptualization of knowledge graphs provides us with guidance as to what might currently "count" as a knowledge graph and also describes capabilities that do not yet exist in current knowledge graphs. From this synthesis, we propose an updated definition along with a set of knowledge graph requirements. We include the requirement that knowledge graphs represent attributable knowledge, thus they need to include information about where the knowledge came from, as opposed to containing "bare statements" with no justification or provenance. We discuss how knowledge graphs as defined are a crucial component for the future of the Web and have great potential for transformational change in data science and domain sciences.

Knowledge graphs provide an opportunity to expand our understanding of how knowledge can be managed on the Web and how that knowledge can be distinguished from more conventional Web-based data publication schemes such as Linked Data [2]. In recent years, knowledge graphs have grown increasingly prominent through commercial and research applications on the Web. Google was one of the first to promote a semantic metadata organizational model described as a "knowledge graph," and many other organizations have since used the term in published research on knowledge management and graph databases. Our purpose with this paper is to survey the evolving notion of a knowledge graph, to describe the general space, and to provide an explicit operational description of a knowledge graph. We begin with a review of recent definitions of knowledge graphs, knowledge graph analysis and construction algorithms, and commercial, research, nonprofit, and government knowledge graphs. These new knowledge graphs do not strictly adhere to original knowledge graph theory [3], but instead have followed a looser, more flexible definition. We present a more descriptive view of current, practical knowledge graphs, and discuss their potential for evolution and impact.

2. Related Work

Rospocher, *et al.* present knowledge graphs as collections of facts about entities, typically derived from structured data sources such as Freebase [4]. They cite a dearth of event representations in current knowledge graphs as a shortcoming - limiting knowledge graphs to encyclopedic items such as birth and death dates - primarily due to the difficulty of obtaining temporal data about entities in a structured manner. Recent surveys, such as those by Hogenboom, *et al.* [5] and Deng, *et al.* [6], provide overviews of numerous methods for event extraction from a variety of sources including social media, news, academic publications, and even images and video, indicating that there is a great interest in finding ways to interpret and include such temporal data in a more structured format. Another review by Nickel *et al.* explores machine learning methods for knowledge graphs, but limits their definition to directed labeled graphs, with the ability to optionally predefine the schema. They also review, but do not take a position on, the use of the closed versus open world assumptions.

van de Riet and Meersman [3], Stokman and de Vries [7], and Zhang [8], present a formal theory of knowledge graphs as a specialization of semantic networks where meaning is expressed as structure, statements are unambiguous, and a limited set of relation types are used. These requirements also minimize redundancy within the knowledge graph, which simplifies analytical operations (including reasoning and queries). Popping explores the use of knowledge graphs, and their challenges at the time, in their use in network text analysis [9]. Following Zhang, Popping defines the knowledge graph as a type of semantic network that uses only a few types of relations, but also asserts that additional knowledge may be added to the graph. Ehrlinger [10] selected some representative definitions that demonstrate the lack of a common core understanding of the concept. Farber, *et al.* [11] and Huang, *et al.* [12] define knowledge graph as being an RDF graph. Paulheim [13] argues that "knowledge graphs are supposed to cover at least a major portion of the domains that exist in the world, and are not supposed to be restricted to only one domain." But while DBpedia or Wikidata are general knowledge graphs and don't focus on a single domain, this should not mean that all knowledge graphs must be general. On the contrary, we believe that knowledge graphs created for specific domains such as Biology can be considered knowledge graphs if they follow the other requirements. More recently, many works report on automatically building knowledge graphs out of textual medical knowledge and medical records [14], [15], [16], [17].

3. A Definition of "Knowledge Graph"

One thing to note is that the knowledge graph platforms that have been reviewed in this paper do not strictly adhere to the definition of knowledge graph that was set out in an de Riet and Meersman [3], Stokman and de Vries [7], and Zhang [8]. Since usage has evolved, it is appropriate to develop a definition that follows how the term is currently used. Implicit in the name, "knowledge graph," is, of course, that a knowledge graph represents *knowledge*, and does so using a *graph* structure. Stokman, de Vries [7], and Zhang [8] posit useful definitions and requirements for knowledge graphs as a starting point:

- Knowledge graph meaning is expressed as structure.
- Knowledge graph statements are unambiguous.
- Knowledge graphs use a limited set of relation types.

Note that the *graph* (in a knowledge graph) is a directed labeled graph. Without direction or labels, it would be impossible to encode any significant meaning into the structure of a graph. In order for knowledge graph statements to be unambiguous, they need to be composed of unambiguous units.

- All identified entities in a knowledge graph, including types and relations, must be identified using global identifiers with unambiguous denotation.

One example of this kind of identifier is the Uniform Resource Identifier (URI), as used in RDF [18].

While the use of "limited set of relation types" proposed by van de Riet *et al.* addressed a specific set of non-decomposable, *essential* relations, in the context of an open world knowledge system this should be taken to mean a core set of *essential classes and relations* that are true regardless of context. For instance, a person is a patient only within the context of a medical encounter. Similarly, in microbiology, there are many proteins that act on other proteins within certain parts of the cell. They are inactive in other parts, so the localization of the protein is context that is needed to understand its behavior.

It is important to consider context as many relations that may seem simple and binary may actually be more complex. This is often seen in DBpedia, where many relationships, like spouses, and children, are expressed as simple triples of the form:

```
John_Lennon spouse Yoko_Ono.
```

However, there is another triple in the graph:

```
John_Lennon spouse Cynthia_Lennon.
```

This suggests that Lennon was married to two people at once. Of course, John divorced Cynthia and married Yoko later on. The *spouse* relationship is actually more complex than modeled in DBpedia, and one reason is because the relation needs information about the time context. The relationship has a beginning, and sometimes an end. If a person has had multiple marriages, that information needs to be added to the relationship, not the person. Expressing these relationships in a vocabulary with limited relations that follow the criteria we introduced above and will revisit below might look more like this:

```
John_Lennon hasRole
  [ a Spouse;
    startTime [ a TimeInstant;
      hasValue "1962-08-23"];
    endTime [ a TimeInstant;
      hasValue "1968-11-08"];
    inRelationTo Cynthia_Lennon];
  [ a Spouse;
    startTime [ a TimeInstant;
      hasValue "1969-03-20"];
    inRelationTo Yoko_Ono].
```

The relations used here allow for elaboration at every level. By limiting knowledge graphs to a set of essential relations, it forces knowledge graph editors to think compositionally, making the graph structure more durable as additional knowledge and context is added. The encoding uses the essential term hasRole and includes the contextual temporal information about when the spouse relationship actually held, thus allowing a statement about John Lennon's spouse relationship to Cynthia DURING a particular context to always be evaluated as true.

In practice, the knowledge graph literature and the practical knowledge graphs we reviewed either aggregate knowledge from many secondary sources and use Natural Language Processing (NLP) extraction when the sources are unstructured text, or use a semantic Extraction Transformation and Load (ETL) process from structured databases [19]. Some knowledge graphs rely on crowdsourcing of their information (including the Google Knowledge Graph), a form of distributed curation. At no point do we see a case where the knowledge does not have a theoretical, citeable source or some other recorded justification. Since knowledge graphs nominally represent knowledge, we argue that some criteria for inclusion of content and its provenance should be encoded in the graph. This is especially true for knowledge graphs gathered from other sources, as the sources themselves must have some justification for publishing their assertions.

– Knowledge graphs must include explicit provenance.

In many cases, the justification for inclusion of assertions appeals to authority, through the citation of the resource the knowledge was extracted from. Authority, at least in scientific research, is only a short cut for validating knowledge, and good knowledge graphs should encode as much justification for their assertions as they can. We consider graphs without provenance, concerning attribution or justification, to be *bare statement graphs*. Bare statement graphs are not true knowledge graphs, according to our definition, since they do not provide a way to confirm that statements are justified or are even believed by their originators; this is a minimal (but not sufficient [20]) criteria for "knowledge" in a knowledge graph.

– Knowledge graphs may include uncertainty assessments.

Some knowledge graphs go further in modeling knowledge by providing uncertainty assessments of the knowledge asserted [21]. This can be useful when dealing with scientific knowledge graphs, where competing hypotheses and theories are known to be true to certain degrees, which may change as new evidence comes to light.

We have therefore identified the following hierarchy of graph types. The basic graph that we build on is a directed labeled graph.

Graph A set of assertions (edges labeled with relations) that are expressed between entities (vertices) where the meaning of the graph is encoded in its structure.

Unambiguous Graph A graph where the relations and entities are unambiguously identified.

Knowledge Graph An Unambiguous Graph with a limited set of relations used to label the edges that encodes the provenance, especially justification and attribution, of the assertions.

All the resources we reviewed are *Graphs*, in the above sense.

4. Knowledge Graph Methods

Corby and Zucker present an abstract knowledge graph querying machine they call KGRAM [22], but do not define knowledge graphs beyond being directed labeled graphs. This work appears to be an abstraction of graph query methods and KGRAM can be viewed as a generalization and extension of the RDF graph query language SPARQL [23]. Wang *et al.* [24] discuss projecting generalized knowledge graphs into hyperplanes, but also only focuses on the labeled directed graph requirement of knowledge graphs. Pujara *et al.* use probabilistic soft logic (PSL) to manage uncertainty in knowledge graphs that have been extracted from uncertain sources [25]. They argue that many current knowledge graphs do not always clearly identify entities, relying instead on labels that can be different due to spelling variations. Their task of "knowledge graph identification" has a goal of identifying a set of true assertions from noisy extractions. They do not claim to manage the provenance of the resulting knowledge graph assertions. Lin *et al.* attempt link prediction for automated knowledge graph construction but only rely on a directed, labeled graph model of knowledge graphs [26]. Hakkani-Tur *et al.* use statistical language understanding to pose structured questions against the Freebase knowledge graph, focusing on improving the extraction of relation detection in the queries [27]. Benedek *et al.* have presented a collaborative knowledge graph construction tool called "Conceptipedia", building off of their "WikiNizer" project [28]. This project uses visual mind mapping techniques and concept similarity analysis to suggest cross-knowledge graph mappings between collaborators. Weiderman and Kritzinger

[29] refer to knowledge graphs as a synonym for concept maps, but do not expand further on the topic, nor do they cite any work in knowledge graphs.

5. A Meta-Knowledge Graph

To organize this paper, we created what we call The Knowledge Graph Catalog (KGC) (http://graphs.whyis.io). The KGC is a meta-knowledge graph that collects metadata about published knowledge graphs and resources that resemble knowledge graphs. It currently describes key features of each graph, the API, publisher, and lifecycle status, and provides a faceted browser of the knowledge resources reported here. We have made an effort to cover as many knowledge graphs as we could find, but readers can contribute to KGC and can suggest knowledge graphs and similar resources, that are not yet in the catalog, through a form on the web site.

5.1. An Ontology of Knowledge Resources

We also developed a knowledge graph catalog ontology (http://graphs.whyis.io/ns) to support the KGC itself. It includes relevant attributes and definitions from here and includes a hierarchy of qualifying knowledge resource types.

6. Knowledge Graphs

We have surveyed 37 potential knowledge graphs so far (including the Knowledge Graph Catalog, or KGC), and have found that 7 of them (also including KGC) fulfill all four requirements of the knowledge graph definition presented in Section 3. We show how each resource fulfills those requirements in Table 1, along with the publisher and production status. They are found in commercial, academic, nonprofit, and government settings. A number of knowledge graphs are experimental or retired, but a significant number are in active production, as shown in Figure 2.

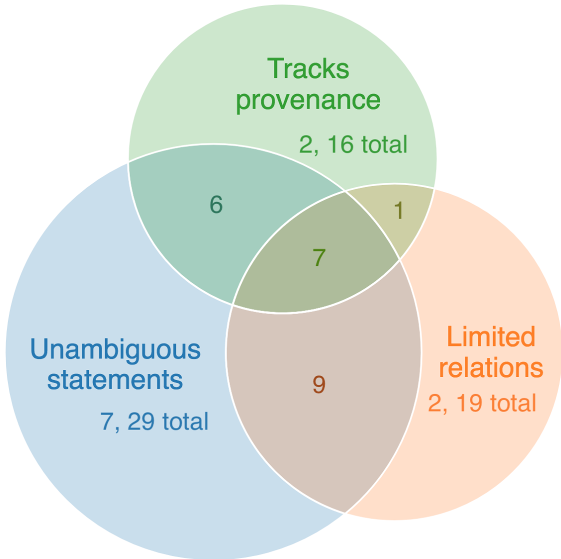

Figure 1. Venn diagram of required features in knowledge graphs. "Meaning as structure" is implemented by all surveyed knowledge graphs, but many graphs do not limit their relations, nor do they track provenance in a meaningful way.

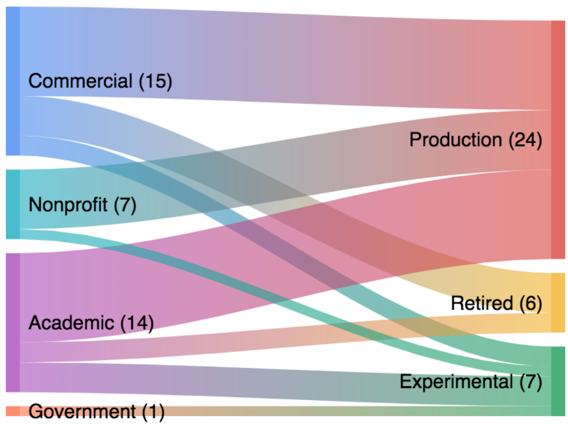

Figure 2. Knowledge graph proportions by publisher and status.

6.1. Academic Knowledge Graphs

The **Gene Ontology** (GO) may be considered more of a knowledge graph than an ontology. It embodies a hierarchy of biological processes, cellular locations, and molecular functions into which a number of genes and proteins have been classified or annotated. These annotations have been curated by domain experts, and the evidence for each is recorded using a GO-specific provenance encoding [30]. Other ontologies in OBO

Foundry also encode knowledge, but do not provide provenance of their assertions. **BioPortal** is a web-based application for accessing and sharing biomedical ontologies. It is the largest such repository, with more than 700 ontologies to date. This set includes ontologies that were developed in OWL, OBO, and other formats, as well as a large number of medical terminologies that the US National Library of Medicine distributes. It supports dereferencing of URIs for whole ontologies and individual terms in the ontologies [31]. The **UniProt Knowledge Base** is an excellent source of manually, expert-verified, Protein data, and provides citations for all assertions made in the graph. Additionally it also provides mappings to other Gene and Protein URI schemes, and relationships with other similar proteins and details about protein interactions. However, a lack of a visual browser on the KG website limits its usability compared to other Knowledge Graph platforms. Furthermore, it also lacks provenance information about the expert that completed the manual verification and Knowledge Graph construction [32]. The **Knowledge Graph Catalog** (KGC) was discussed in Section 5.

6.2. Commercial Knowledge Graphs

Google introduced its **Knowledge Graph project** in 2012, and has used it to improve query result relevancy and their overall search experience. They have leveraged existing knowledge graphs, such as DBpedia and Freebase, and also have opened up the process of contributing to the graph by ingesting RDFa and microdata formats from the Web pages they index, based on the vocabularies published by schema.org [1]. The **Knowledge Vault**, a research project funded by Google, handles knowledge graph uncertainty as a result of automated fact extraction from Web pages, and attempts to fuse data from multiple sources into a singular knowledge graph [33].

6.3. Nonprofit Knowledge Graphs

The **Nexus Knowledge Graph** is a schema-driven knowledge graph that uses the W3C PROV ontology to manage provenance about contributed knowledge. It focuses on neuroscience, but developers claim the core technology will apply to other domains [34].

7. Other Graph Resources

While one could argue that the following resources are knowledge graphs, they do not fulfill the complete definition, even if they contain some of the requirements. All of these knowledge graphs express meaning as structure, but they all fail to provide one or more of unambiguous identifiers, limited sets of relations, or knowledge provenance.

7.1. Academic Graph Resources

BabelNet is a multilingual knowledge graph that attempts automated entity disambiguation across its languages, and provides an integration of Wikipedia and WordNet [35]. **Chemical Entities of Biological Interest** (ChEBI) relies on hand-annotation and affirmation of concepts, which ensures that the data contained within is accurate according to domain scientists, at the expense of taking longer to provide updates to the knowledge graph. One of the primary focuses of the original effort was to unite a variety of molecular representation and nomenclature standards utilizing unambiguous identifiers. This gives credibility and re-usability as a resource in the relevant chemical and biochemical communities. Previous versions overloaded some of the relationship terms, which introduced difficulty in representing the information in an ontology graph format, but more recent versions have worked to resolve these issues [36–38]. **ConceptNet** is a knowledge graph of things people know, and computers should know, expressed in various natural languages. ConceptNet uses PostgreSQL as database [39]. **DBPedia** is a large-scale transformation of Wikipedia into a knowledge graph. Since it is curated from crowdsourced wikipedia infoboxes, it contains an unlimited set of relations. It also relies on Wikipedia for the provenance of any changes to the knowledge graph, since it is generated from there [40]. **Elementary** (later, DeepDive) is a framework for developing knowledge graphs using natural language processing algorithms to extract and infer new knowledge [41]. **PROSPERA** is another knowledge extraction tool that attempts scalable extraction with high precision and recall [42]. **Read the Web/Never Ending Learning** uses semi-supervised natural language processing techniques to build a knowledge graph against a set of entity types. In the web interface, each assertion is linked back to one or more web pages it was extracted from [43]. **ReVerb** is a knowledge extraction tool that extracts binary relationships from text without the need for a pre-defined ontology. This has the benefit of completeness, but allows for ambiguous expressions of knowledge within its output [44]. **GeoLink** presents an extensive collection of Geography-research related data. Additionally, the application developed around the graph (http://demo.geolink.org) is well implemented, and provides the ability to map individual measurements to a map; a particularly useful feature for a Geography knowledge graph. However, it appears that some of the data in the graph is not fully linked; for example, certain physical measurements may not have associated scientists, or vice versa. Additionally, GeoLink does not use a system of Nanopublications, and thus, does not preserve provenance of some of the nodes in the graph [45]. The **Neurocommons** project attempted to represent all biological knowledge relating to neuroscience research as a common knowledge graph by integrating a number of biological databases. The project was still in development when funding ended [46]. The **XLore** system claims to be a fully bilingual (Chinese and English) knowledge graph that focuses on extracting *subClassOf* and *instanceOf* relations from free text [47]. **Bio2RDF** was originally designed as a meta-search over a variety of existing biomedical vocabularies and ontologies. While it provides good unification of synonyms over a variety of source ontologies, keeping the sources separated is also important to this effort, and so there is less overall integration, and the representation is a direct mapping of the source data into RDF [48, 49]. **YAGO** (Yet Another Great Ontology) is considered by some researchers to be a knowledge graph, although it originated as a large, general-purpose ontology. While they aggregate knowledge from many sources, there are no published descriptions of whether or how provenance is tracked in YAGO [50].

7.2. Commercial Graph Resources

Cyc, started in 1987, is the world's longest-running artificial inteligence project. Cyc is principally based in predicate logic and provides default values for logic (inferred values can be overridden by explicit assertions). Cyc has over 15,000 predicates, which precludes it from having a limited set of relations [51]. **OpenCyc**, which was discontinued in 2017, was an RDF-based open source subset of Cyc [52]. **Freebase** is a knowledge graph of over 3B facts and 58M topics (Freebase.com web site, April 2016) that was open to public access and curation and formed the basis for the Google Knowledge Graph. It has since retired and is available as an RDF download [53]. IOS Press has developed a linked data portal for its publication metadata using the BIBO ontology. Since it is the authority for publication metadata, it does not directly encode provenance of this metadata itself [54]. **Linked Life Data** is constructed from a number of biomedical databases, using its own internal vocabulary for a data model [55]. **Open Knowledge Graph** was a project that aggregated the output of the **SEKI@home** project. SEKI@home is a crowd-sourced knowledge graph that aggregates from multiple sources, maintaining entity-level provenance using the PROV Ontology [56]. **Probase** is a automatically generated taxonomy of classes and instances automatically extracted from the web [57]. The **Suggested Upper Merged Ontology** (SUMO) attempts to provide an "all in one" graph of knowledge, and includes partial mappings to WordNet and DBPedia. It uses the knowledge interchange format, and only provides partial serializations in OWL [58]. The **Bing Entity Search** API is the first public API to be released from Microsoft's **Satori** knowledge graph project. It focuses on resolving entities and providing information about them to API users [59]. The **Face-**

book Graph API provides access to the "social graph" expressed in the Facebook social networking platform. It encodes social knowledge, including, for users who have the right permissions, access to social interactions via the Facebook platform [60]. The popular **Wolfram Alpha** natural language, mathematics engine utilizes a version of a conventional ontology, implemented using symbolic programming. The Wolfram Alpha API is planned to expose some of the underlying Ontology to developers in a future release [61]. **Walmart** has funded research into a knowledge resource as well, extracting structured knowledge from Wikipedia. The effort seems to be similar to DBpedia, but has not yet produced any public output [62]. **Upper Mapping and Binding Exchange Layer** (UMBEL) is designed to be a source of entity content. It is designed to provide a coherent ontology for aligning specific entity knowledge within a broader context. UMBEL used to be distributed and maintained by Structured Dynamics LLC. The UMBEL Ontology has not been updated since the company ceased operations in 2016 [63]. **Unigraph** attempts to aggregate knowledge graphs from across the web, but it is unclear from the documentation how entities and relations are identified [64].

7.3. Government Graph Resources

The United States Geological Survey **Geographic Names Information System** (GNIS) has an experimental linked data representation that uses GeoSPARQL to provide geospatial indexing of geographic features [65].

7.4. Nonprofit Graph Resources

Ontobee has a web interface for querying and visualizing the details and hierarchy of a specific ontology term. It is able to dereference a single ontology term URI, and then display the HTML information on a browser. Statistics and other detailed information are generated and displayed. A SPARQL web interface is provided for custom queries [66]. The **Environment Ontology** (ENVO) is an ontology of classes relating to environmental research. It is an OWL ontology and part of the OBO Foundry, and is available through Ontobee [67]. The **ESKG Knowledge Graph** provides a thorough interface to NASA's archived Earth-Science studies and data. Additionally, mathematical units and concepts are expressed using unambiguous identifiers. However, there is no user interface provided and the knowledge graph is generated automatically and may not be error-free [68]. **Wikidata** is a knowledge graph developed by the Wikimedia Foundation as an effort to provide structured data to Wikipedia and other efforts. It has developed a language-independent identifier system for entities, and all information is available as an RDF graph. It encourages and allows for references on a per-assertion basis to provide evidence to support it. It also tracks who creates and modifies facts in the knowledge graph [69].

8. Future Potential

Usually, knowledge graphs are not distinguished from bare statement graphs, in that they do not encode or publish the epistemology[2] of knowledge asserted in the graph. We see this as troubling because it does not *privilege* knowledge; in most existing knowledge graphs, supported and unsupported assertions are given equal weight. Moving forward, there is an opportunity to leverage existing vocabularies, including the Provenance Ontology (PROV-O) [70], and the Nanopublications Framework [71], to improve the clarity, transparency, and utility of knowledge graphs. A nanopublication is a set of RDF graphs: an *assertion graph* (the knowledge), a *provenance graph* (the justification), and an *attribution graph* (the believer). While justified true belief is not sufficient for knowledge, most other proposals, including a causal linkage between the justification, assertion, and believer, are well-supported within provenance vocabularies. Added to a knowledge graph, provenance graphs can expand to provide room for whatever epistemic criteria is desired.

There is an interesting overlap between what is considered a “knowledge graph” and what is an

[2]Epistemology defines why something is known.

Table 1
A breakdown of the features of the reviewed knowledge graphs.

Status	Publisher	Name	Structured Meaning	Unambiguous	Tracks provenance	Limited relations
Experimental	Carnegie Mellon University	Read the Web	Yes	No	Yes	No
	Max Planck Institute for Informatics	PROSPERA	Yes	No	No	No
	University of Washington	ReVerb	Yes	No	No	No
	Google	Knowledge Vault	Yes	Yes	Yes	Yes
	Microsoft	Probase	Yes	No	No	No
	Walmart	Walmart Lab's Social Genome	Yes	No	No	No
	United States Geological Survey	Geographic Names Information System	Yes	Yes	No	No
	Blue Brain Project	Nexus KnowledgeGraph	Yes	Yes	Yes	Yes
Production	Luminoso Technologies, Inc.	ConceptNet	Yes	Yes	No	Yes
	European Bioinformatics Institute	Chemical Entities of Biological Interest (ChEBI)	Yes	Yes	Partial	Yes
		UniProt KB	Yes	Yes	Yes	Yes
	Laval University	BIO2RDF	Yes	Yes	Partial	
	Leipzig University	DBpedia	Yes	Yes	Yes	No
	Max Planck Institute for Informatics	Yet Another Great Ontology	Yes	Yes	No	No
	National Science Foundation	EarthCube GeoLink	No	No	No	Yes
	OBO Foundry	Gene Ontology	Yes	Yes	Yes	Yes
	National Center for Biomedical Ontology	BioPortal	Yes	Yes	Yes	Yes
	Rensselaer Polytechnic Institute	Knowledge Graph Catalog	Yes	Yes	Yes	Yes
	Sapienza University of Rome	BabelNet	Yes	Yes	No	No
	Tsinghua University	XLore	Yes	Yes	No	No
	Articulate Software	Suggested Upper Merged Ontology	Yes	Partial	No	Partial
	Cycorp	Cyc	Yes	Yes	Yes	No
	Facebook	Facebook Graph API	Yes	No	No	Yes
	Google	Google Knowledge Graph	Yes	Yes	Yes	Yes
	INGENIOSITY LTD	Unigraph	Yes	No	Yes	No
	IOS Press	LD Connect	Yes	Yes	No	Yes
	Microsoft	Bing Entity Search API	Yes	No	No	No
	Ontotext	Linked Life Data	Yes	Yes	Yes	No
	Thompson Reuters	Thomson Reuters Knowledge Graph Feed	Yes	Yes	Don't Know	Don't Know
	Wolfram Alpha	Wolfram Alpha Internal Knowledge Graph	Yes	No	Yes	Yes
	Earth Science Information Partners	Earth Science Knowledge Graph	Yes	Yes	No	Yes
	EDM Council	Financial Industry Business Ontology (FIBO)	Yes	Yes	No	Yes
	OBO Foundry	Environment Ontology	Yes	Yes	No	Yes
		Ontobee	Yes	Yes	No	Yes
	Wikimedia Foundation	Wikidata	Yes	Yes	Yes	No
Retired	Science Commons	Neurocommons	Yes	Yes	No	Yes
	Stanford University	Elementary/DeepDive	Yes	No	No	No
	Cycorp	OpenCyc	Yes	Yes	No	No
	Google	Freebase	Yes	Yes	Yes	No
		Open Knowledge Graph	Yes	Yes	Yes	No
Unmaintained	Structured Dynamics LLC	Upper Mapping and Binding Exchange Layer	Yes	Yes	No	Yes

ontology. The most commonly accepted definition of an ontology is "an explicit specification of a conceptualization" [72]. To a large degree, knowledge graphs conform to this definition, but generally ontologies tend to talk about generalities (classes, properties, and roles) with less focus on inclusion of content about specific instances. For example, most ontologies that include content related to descriptions of world landmarks would have descriptions of the landmark class and its related properties but would typically not include a mention of the Eiffel Tower, but a knowledge graph that covers the domain of Parisian landmarks, would. Conversely, knowledge graph approaches can be used to improve the credibility of ontologies by encoding the epistemology of the statements in the ontology.

9. Conclusion

Knowledge graphs are an increasingly critical component of the Semantic Web and serve as information hubs for general use as well as for domain-specific applications. Most knowledge graphs seek to aggregate knowledge from third party sources, whether from external databases, from data aggregated though crawling the Web, or through the application of entity and relationship extraction methods. Knowledge graphs are not simply aggregations of RDF or linked data, but critically provide time-invariant information about entities of general interest. Their structures tend to be focused on a limited set of relations adhering to a coherent knowledge model, setting them apart from the linked data cloud in general, which usually has relied on the open framework of the Semantic Web to accommodate a completely free-form use of vocabularies and ontologies. Although some knowledge graphs track the provenance of their content, rigorous provenance is by no means a universal characteristic. We argue that knowledge graphs should prioritize the epistemology of the knowledge it contains – how we know what we know – and that Nanopublications are a suitable framework in which to do so. Semantic publishing that does not provide a level of statement epistemology can be considered "Bare Statement" graphs. Since so many knowledge graphs are curated from third parties, and because of the nature of publishing on the Web (*Anyone* can say *Anything* about *Any* subject), as knowledge graphs increase in popularity it will become critical to avoid use of such "Bare Statement" graphs. We also show that, while most knowledge resources can be considered *graphs*, in the sense defined here, not all can be considered *unambiguous graphs*, in that they do not use unambiguous identifiers and do not use a limited set of relations. We hope that these definitions help provide a means for setting expectations for knowledge resources, and to help guide and refine the scope of knowledge graph research.

Acknowledgements

This work was funded by National Spectrum Consortium's Dynamic Spectrum Access Policy Development project, NIEHS Award 0255-0236-4609 / 1U2CES026555-01, NSF Award OAC-1640840 IBM Research AI Horizons Network, and by the Gates Foundation through HBGDki.

References

[1] A. Singhal, Introducing the knowledge graph: things, not strings, *Official Google Blog, May* (2012), Accessed: 2016-04-11. https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.

[2] C. Bizer, T. Heath and T. Berners-Lee, Linked data-the story so far, *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (2009), 205–227.

[3] R. van de Riet and R. Meersman, Knowledge Graphs, in: *Linguistic Instruments in Knowledge Engineering: Proceedings of the 1991 Workshop on Linguistic Instruments in Knowledge Engineering, Tilburg, the Netherlands, 17-18 January 1991*, North-Holland, 1992, p. 97.

[4] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger and T. Bogaard, Building event-centric knowledge graphs from news, *Web Semantics: Science, Services and Agents on the World Wide Web* (2016), ISSN 1570-8268. doi:10.1016/j.websem.2015.12.004. http://www.sciencedirect.com/science/article/pii/S1570826815001456.

[5] F. Hogenboom, F. Frasincar, U. Kaymak, F. de Jong and E. Caron, A Survey of event extraction methods from text for decision support systems, *Decision Support Systems* (2016), ISSN 0167-9236. doi:10.1016/j.dss.2016.02.006. http://www.sciencedirect.com/science/article/pii/S0167923616300173.

[6] J. Deng, F. Qiao, H. Li, X. Zhang and H. Wang, An Overview of Event Extraction from Twitter, in: *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2015 International Conference on*, 2015, pp. 251–256. doi:10.1109/CyberC.2015.24.

[7] F.N. Stokman and P.H. de Vries, Structuring Knowledge in a Graph, in: *Human-Computer Interaction*, Springer Science + Business Media, 1988, pp. 186–206. doi:10.1007/978-3-642-73402-1_12. http://dx.doi.org/10.1007/978-3-642-73402-1_12.

[8] L. Zhang, *Knowledge graph theory and structural parsing*, Twente University Press, 2002.

[9] R. Popping, Knowledge Graphs and Network Text Analysis, *Social Science Information* **42**(1) (2003), 91–106. doi:10.1177/0539018403042001798. http://dx.doi.org/10.1177/0539018403042001798.

[10] L. Ehrlinger and W. Wöß, Towards a Definition of Knowledge Graphs., in: *SEMANTiCS (Posters, Demos, SuCCESS)*, 2016.

[11] M. Färber, F. Bartscherer, C. Menne and A. Rettinger, Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago, *Semantic Web* (2016), 1–53.

[12] Z. Huang, J. Yang, F. van Harmelen and Q. Hu, Constructing disease-centric knowledge graphs: a case study for depression (short version), in: *Conference on Artificial Intelligence in Medicine in Europe*, Springer, 2017, pp. 48–52.

[13] H. Paulheim, Knowledge graph refinement: A survey of approaches and evaluation methods, *Semantic web* **8**(3) (2017), 489–508.

[14] L. Shi, S. Li, X. Yang, J. Qi, G. Pan and B. Zhou, Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services, *BioMed Research International* **2017** (2017).

[15] A. Lamurias, J.D. Ferreira, L.A. Clarke and F.M. Couto, generating a Tolerogenic cell Therapy Knowledge graph from literature, *Frontiers in immunology* **8** (2017), 1656.

[16] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng and D. Sontag, Learning a health knowledge graph from electronic medical records, *Scientific reports* **7**(1) (2017), 5994.

[17] P. Ernst, A. Siu and G. Weikum, KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences, *BMC bioinformatics* **16**(1) (2015), 157.

[18] R. Cyganiak, D. Wood and M. Lanthaler, RDF 1.1 concepts and abstract syntax, *W3C Recommendation. Feb* (2014).

[19] J.P. McCusker, J.A. Phillips, A. Beltrán, A. Finkelstein and M. Krauthammer, Semantic web data warehousing for caGrid, *BMC Bioinformatics* **10**(Suppl 10) (2009), 2. doi:10.1186/1471-2105-10-s10-s2. http://dx.doi.org/10.1186/1471-2105-10-S10-S2.

[20] E.L. Gettier, Is Justified True Belief Knowledge?, *Analysis* **23**(6) (1963), 121–123. doi:10.1093/analys/23.6.121. http://dx.doi.org/10.1093/analys/23.6.121.

[21] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun and W. Zhang, Knowledge vault, in: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, Association for Computing Machinery (ACM), 2014. doi:10.1145/2623330.2623623. http://dx.doi.org/10.1145/2623330.2623623.

[22] O. Corby and C.F. Zucker, The KGRAM Abstract Machine for Knowledge Graph Querying, in: *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Institute of Electrical & Electronics Engineers (IEEE), 2010. doi:10.1109/wi-iat.2010.144. http://dx.doi.org/10.1109/WI-IAT.2010.144.

[23] S. Harris, A. Seaborne and E. Prud'hommeaux, SPARQL 1.1 query language, *W3C Recommendation* **21** (2013).

[24] Z. Wang, J. Zhang, J. Feng and Z. Chen, Knowledge Graph Embedding by Translating on Hyperplanes, in: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[25] J. Pujara, H. Miao, L. Getoor and W. Cohen, Knowledge Graph Identification, in: *Lecture Notes in Computer Science*, Springer Science + Business Media, 2013, pp. 542–557. doi:10.1007/978-3-642-41335-3_34. http://dx.doi.org/10.1007/978-3-642-41335-3_34.

[26] Y. Lin, Z. Liu, M. Sun, Y. Liu and X. Zhu, Learning Entity and Relation Embeddings for Knowledge Graph Completion., in: *AAAI*, 2015, pp. 2181–2187.

[27] D. Hakkani-Tur, L. Heck and G. Tur, Using a knowledge graph and query click logs for unsupervised learning of relation detection, in: *2013 IEEE International Conference on Acoustics Speech and Signal Processing*, Institute of Electrical & Electronics Engineers (IEEE), 2013. doi:10.1109/icassp.2013.6639289. http://dx.doi.org/10.1109/ICASSP.2013.6639289.

[28] A. Benedek, C. Goodman and G. Lajos, The 'Conceptipedia' of Visual Semantic Wikinizers: A Reference Model For Collaborative Conceptualization, *INNODOCT/13 "New changes in technology and innovation"*, 109.

[29] M. Weideman and W. Kritzinger, Concept Mapping vs. Web Page Hyperlinks As an Information Retrieval Interface: Preferences of Postgraduate Culturally Diverse Learners, in: *Proceedings of the 2003 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on Enablement Through Technology*, SAICSIT '03, South African Institute for Computer Scientists and Informa-

tion Technologists, Republic of South Africa, 2003, pp. 69–82. ISBN 1-58113-774-5. http://dl.acm.org/citation.cfm?id=954014.954022.

[30] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G. Sherlock, Gene Ontology: tool for the unification of biology, *Nature Genetics* **25**(1) (2000), 25–29. doi:10.1038/75556. https://doi.org/10.1038/75556.

[31] S. Manuel, A.P. R., M.M. A. and N.N. F., BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF, *Semantic Web* **4**(3) (2013), 277–284, ISSN 1570-0844. doi:10.3233/SW-2012-0086. http://doi.org/10.3233/SW-2012-0086.

[32] UniProt: the universal protein knowledgebase, *Nucleic Acids Research* **45**(D1) (2016), 158–169. doi:10.1093/nar/gkw1099. https://doi.org/10.1093/nar/gkw1099.

[33] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun and W. Zhang, Knowledge vault, in: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, ACM Press, 2014. doi:10.1145/2623330.2623623. https://doi.org/10.1145/2623330.2623623.

[34] Blue Brain Nexus, Last accessed 6/27/2018. https://bbp-nexus.epfl.ch/staging/home.

[35] R. Navigli and S.P. Ponzetto, BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, *Artificial Intelligence* **193** (2012), 217–250. doi:10.1016/j.artint.2012.07.001. https://doi.org/10.1016/j.artint.2012.07.001.

[36] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj and M. Ashburner, ChEBI: a database and ontology for chemical entities of biological interest, *Nucleic Acids Research* **36**(Database) (2007), 344–350. doi:10.1093/nar/gkm791. https://doi.org/10.1093/nar/gkm791.

[37] P. de Matos, R. Alcántara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner and C. Steinbeck, Chemical Entities of Biological Interest: an update, *Nucleic Acids Research* **38**(suppl_1) (2009), 249–254. doi:10.1093/nar/gkp886. https://doi.org/10.1093/nar/gkp886.

[38] J. Hastings, P. de Matos, A. Dekker, M. Ennis, B. Harsha, N. Kale, V. Muthukrishnan, G. Owen, S. Turner, M. Williams and C. Steinbeck, The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013, *Nucleic Acids Research* **41**(D1) (2012), 456–463. doi:10.1093/nar/gks1146. https://doi.org/10.1093/nar/gks1146.

[39] R. Speer and C. Havasi, ConceptNet 5: A Large Semantic Network for Relational Knowledge, in: *The People's Web Meets NLP*, Springer Berlin Heidelberg, 2013, pp. 161–176. doi:10.1007/978-3-642-35085-6_6. https://doi.org/10.1007/978-3-642-35085-6_6.

[40] L. Jens, I. Robert, J. Max, J. Anja, K. Dimitris, M.P. N., H. Sebastian, M. Mohamed, van Kleef Patrick, A. Soren and et al., DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia, *Semantic Web* **6**(2) (2015), 167–195, ISSN 1570-0844. doi:10.3233/SW-140134. http://doi.org/10.3233/SW-140134.

[41] F. Niu, C. Zhang, C. Ré and J. Shavlik, Elementary, *International Journal on Semantic Web and Information Systems* **8**(3) (2012), 42–73. doi:10.4018/jswis.2012070103. https://doi.org/10.4018/jswis.2012070103.

[42] N. Nakashole, M. Theobald and G. Weikum, Scalable knowledge harvesting with high precision and high recall, in: *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*, ACM Press, 2011. doi:10.1145/1935826.1935869. https://doi.org/10.1145/1935826.1935869.

[43] T. Mitchell, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios, A. Ritter, M. Samadi, B. Settles, W. Cohen, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov, M. Greaves, J. Welling, E. Hruschka, P. Talukdar, B. Yang, J. Betteridge, A. Carlson, B. Dalvi and M. Gardner, Never-ending learning, *Communications of the ACM* **61**(5) (2018), 103–115. doi:10.1145/3191513. https://doi.org/10.1145/3191513.

[44] A. Fader, S. Soderland and O. Etzioni, Identifying Relations for Open Information Extraction, in: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 1535–1545. ISBN 978-1-937284-11-4. http://dl.acm.org/citation.cfm?id=2145432.2145596.

[45] A. Krisnadhi, Y. Hu, K. Janowicz, P. Hitzler, R. Arko, S. Carbotte, C. Chandler, M. Cheatham, D. Fils, T. Finin, P. Ji, M. Jones, N. Karima, K. Lehnert, A. Mickle, T. Narock, M. O'Brien, L. Raymond, A. Shepherd, M. Schildhauer and P. Wiebe, The GeoLink Modular Oceanography Ontology, in: *The Semantic Web - ISWC 2015*, Springer International Publishing, 2015, pp. 301–309. doi:10.1007/978-3-319-25010-6_19. https://doi.org/10.1007/978-3-319-25010-6_19.

[46] A. Ruttenberg, J.A. Rees, M. Samwald and M.S. Marshall, Life sciences on the Semantic Web: the Neurocommons and beyond, *Briefings in Bioinformatics* **10**(2) (2009), 193–204. doi:10.1093/bib/bbp004. https://doi.org/10.1093/bib/bbp004.

[47] Z. Wang, J. Li, Z. Wang, S. Li, M. Li, D. Zhang, Y. Shi, Y. Liu, P. Zhang and J. Tang, Xlore: A large-scale english-chinese bilingual knowledge graph, in: *Proceedings of the 2013th International Conference on Posters & Demonstrations Track-Volume 1035*, CEUR-WS. org, 2013, pp. 121–124.

[48] F. Belleau, M.-A. Nolin, N. Tourigny, P. Rigault and J. Morissette, Bio2RDF: Towards a mashup to build bioinformatics knowledge systems, *Journal of Biomedical Informatics* **41**(5) (2008), 706–716. doi:10.1016/j.jbi.2008.03.004. https://doi.org/10.1016/j.jbi.2008.03.004.

[49] A. Callahan, J. Cruz-Toledo, P. Ansell and M. Dumontier, Bio2RDF Release 2: Improved Coverage, Interoperability and Provenance of Life Science Linked Data, in: *The Semantic Web: Semantics and Big Data*, Springer Berlin Heidelberg, 2013, pp. 200–212. doi:10.1007/978-3-642-38288-8_14. https://doi.org/10.1007/978-3-642-38288-8_14.

[50] F.M. Suchanek, G. Kasneci and G. Weikum, Yago, in: *Proceedings of the 16th international conference on World Wide Web - WWW 2007*, ACM Press, 2007. doi:10.1145/1242572.1242667. https://doi.org/10.1145/1242572.1242667.

[51] D.B. Lenat, CYC: A large-scale investment in knowledge infrastructure, *Communications of the ACM* **38**(11) (1995), 33–38.

[52] M.A. Sicilia, E. García, S. Sánchez and E. Rodríguez, On integrating learning object metadata inside the OpenCyc knowledge base, in: *Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on*, IEEE, 2004, pp. 900–901.

[53] K. Bollacker, C. Evans, P. Paritosh, T. Sturge and J. Taylor, Freebase, in: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data - SIGMOD '08*, ACM Press, 2008. doi:10.1145/1376616.1376746. https://doi.org/10.1145/1376616.1376746.

[54] Welcome To LD Connect, Our Linked Data Portal. http://ld.iospress.nl/.

[55] V. Momtchev, D. Peychev, T. Primov and G. Georgiev, Expanding the pathway and interaction knowledge in linked life data, in: *In Proc. of International Semantic Web Challenge*, 2009.

[56] T. Steiner and S. Mirea, SEKI@ home, or Crowdsourcing an Open Knowledge Graph, in: *Proceedings of the First International Workshop on Knowledge Extraction and Consolidation from Social Media (KECSM2012)*, 2012, p. 7.

[57] W. Wu, H. Li, H. Wang and K.Q. Zhu, Probase, in: *Proceedings of the 2012 international conference on Management of Data - SIGMOD '12*, ACM Press, 2012. doi:10.1145/2213836.2213891. https://doi.org/10.1145/2213836.2213891.

[58] A. Pease, I. Niles and J. Li, The suggested upper merged ontology: A large ontology for the semantic web and its applications, in: *Working notes of the AAAI-2002 workshop on ontologies and the semantic web*, Vol. 28, 2002, pp. 7–10.

[59] R. Qian, Understand Your World with Bing, bing search blog, 2013, Accessed: 2016-04-11. http://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/.

[60] W. Jesse and T. Paul, Facebook Linked Data via the Graph API, *Semantic Web* **4**(3) (2013), 245–250, ISSN 1570-0844. doi:10.3233/SW-2012-0078. http://doi.org/10.3233/SW-2012-0078.

[61] S. Wolfram, The Reform of Mathematical Notation, Computation, Mathematical Notation, and Linguistics, *Inc., Mathematica, Version* **8** (2013), 23.

[62] O. Deshpande, D.S. Lamba, M. Tourn, S. Das, S. Subramaniam, A. Rajaraman, V. Harinarayan and A. Doan, Building, maintaining, and using knowledge bases, in: *Proceedings of the 2013 international conference on Management of data - SIGMOD '13*, ACM Press, 2013. doi:10.1145/2463676.2465297. https://doi.org/10.1145/2463676.2465297.

[63] M.K. Bergman and F. Giasson, UMBEL ontology, Technical Report, Structured Dynamics, 2008. http://umbel.org.

[64] The World's Knowledge Graph. https://unigraph.io/.

[65] B. Regalia, K. Janowicz, G. Mai, D. Varanka and E.L. Usery, GNIS-LD: Serving and Visualizing the Geographic Names Information System Gazetteer as Linked Data, in: *The Semantic Web*, Springer International Publishing, 2018, pp. 528–540. doi:10.1007/978-3-319-93417-4_34. https://doi.org/10.1007/978-3-319-93417-4_34.

[66] E. Ong, Z. Xiang, B. Zhao, Y. Liu, Y. Lin, J. Zheng, C. Mungall, M. Courtot, A. Ruttenberg and Y. He, Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration, *Nucleic Acids Research* **45**(D1) (2016), 347–352. doi:10.1093/nar/gkw918. https://doi.org/10.1093/nar/gkw918.

[67] P.L. Buttigieg, E. Pafilis, S.E. Lewis, M.P. Schildhauer, R.L. Walls and C.J. Mungall, The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation, *Journal of Biomedical Semantics* **7**(1) (2016). doi:10.1186/s13326-016-0097-6. https://doi.org/10.1186/s13326-016-0097-6.

[68] McGibbney, L. J., Jiang and A. B., ESIP's Earth Science Knowledge Graph (ESKG) Testbed Project: An Automatic Approach to Building Interdisciplinary Earth Science Knowledge Graphs to Improve Data Discovery, 2017. http://adsabs.harvard.edu/abs/2017AGUFMIN33C0131M.

[69] D. Vrandečić and M. Kr¨otzsch, Wikidata, *Communications of the ACM* **57**(10) (2014), 78–85. doi:10.1145/2629489. https://doi.org/10.1145/2629489.

[70] L. Moreau, P. Groth, J. Cheney, T. Lebo and S. Miles, The rationale of PROV, *Web Semantics: Science Services and Agents on the World Wide Web* **35** (2015), 235–257. doi:10.1016/j.websem.2015.04.001. http://dx.doi.org/10.1016/j.websem.2015.04.001.

[71] P. Groth, A. Gibson and J. Velterop, The anatomy of a nanopublication, *Information Services and Use* **30**(1–2) (2010), 51–56.

[72] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition* **5**(2) (1993), 199–220. doi:10.1006/knac.1993.1008. http://dx.doi.org/10.1006/knac.1993.1008.