The online encyclopedia Wikipedia is being supplemented by user-edited structured data, available for free to anyone.

BY DENNY VRANDEČIĆ AND MARKUS KRÖTZSCH

Wikidata: A Free Collaborative Knowledge Base

UNNOTICED BY MOST of its readers, Wikipedia is currently undergoing dramatic changes, as its sister project Wikidata introduces a new multilingual 'Wikipedia for data' to manage the factual information of the popular online encyclopedia. With Wikipedia's data becoming cleaned and integrated in a single location, opportunities arise for many new applications.

Initially conceived as a mostly text-based resource, Wikipedia [1] has been collecting increasing amounts of structured data: numbers, dates, coordinates, and many types of relationships from family trees to the taxonomy of species. This data has become a resource of enormous value, with potential applications across all areas of science, technology, and culture. This development is hardly surprising given that Wikipedia is driven by the general vision of 'a world in which every single human being can freely share in the sum of all knowledge'. There can be no question today that this sum must include data that can be searched, analyzed, and reused.

It may thus be surprising that Wikipedia does not provide direct access to most of this data, neither through query services nor through downloadable data exports. Actual uses of the data are rare and often restricted to very specific pieces of information, such as the geo-tags of Wikipedia articles used in Google Maps. The reason for this striking gap between vision and reality is that Wikipedia's data is buried within 30 million Wikipedia articles in 287 languages, from where it is very difficult to extract.

This situation is unfortunate for anyone who wants to make use of the data, but it is also an increasing threat to Wikipedia's main goal of providing up-to-date and accurate encyclopedic knowledge. The same information often appears in articles in many languages and on many articles within a single language. Population numbers for Rome, for example, can be found in the English and Italian article about Rome, but also in the English article *Cities in Italy*. All of these numbers are different.

The goal of Wikidata is to overcome these problems by creating new ways for Wikipedia to manage its data on a global scale. The result of these ongoing efforts can be seen at wikidata.org. The following essential design decisions characterize the approach taken by Wikidata. We will have a closer look at some of these points later.

Open Editing. Like Wikipedia, Wikidata allows every user of the site to extend and edit the stored information, even without creating an account. A form-based interface makes editing very easy.

Community Control. Not only the actual data but also the schema of the data is controlled by the contributor community. Contributors edit the population number of Rome, but they also decide that there is such a number in the first place.

Plurality. It would be naive to expect global agreement on the 'true' data, since many facts are disputed or simply uncertain. Wikidata allows conflicting data to co-exist and provides mechanisms to organize this plurality.

Secondary Data. Wikidata gathers facts published in primary sources, together with references to these sources. There is no 'true population of Rome', but a 'population of Rome as published by the city of Rome in 2011'.

Multilingual Data. Most data is not tied to one language: numbers, dates, and coordinates have universal meaning; labels like *Rome* and *population* are translated into many languages. Wikidata is multi-lingual by design. While Wikipedia has independent editions for each language, there is only one Wikidata site.

Easy Access. Wikidata's goal is to allow data to be used both in Wikipedia and in external applications. Data is exported through Web services in several formats, including JSON and RDF. Data is published under legal terms that allow the widest possible reuse.

Continuous Evolution. In the best tradition of Wikipedia, Wikidata grows with its community and tasks. Instead of developing a perfect system that is presented to the world in a couple of years, new features are deployed incrementally and as early as possible.

These properties characterize Wikidata as a specific kind of curated database [8].

Data in Wikipedia: The Story So Far

The value of Wikipedia's data has long been obvious, and many attempts have been made to use it. The approach of Wikidata is to crowdsource data acquisition, allowing a global community to edit data. This extends the traditional *wiki* approach of allowing users to edit a website (*wiki* is a Hawaiian word for *fast*; Ward Cunning-

ham, who created the first wiki in 1995, used it to emphasize that his website could be changed quickly [17]).

The most popular such system is Semantic MediaWiki (SMW) [15], which extends MediaWiki, the software used to run Wikipedia [2], with data management capabilities. SMW was originally proposed for Wikipedia, but soon was used on hundreds of other websites instead. In contrast to Wikidata, SMW manages data as part of its textual content. This hinders the creation of a multilingual, single knowledge base supporting all Wikimedia projects. Moreover, the data model of Wikidata (discussed below) is more elaborate than that of SMW, allowing users to capture more complex information. In spite of these differences, SMW has had a great influence on Wikidata, and the two projects are sharing code for common tasks.

Other examples of free knowledge base projects are OpenCyc and Freebase. OpenCyc is the free part of Cyc [16], which aims for a much more comprehensive and expressive representation of knowledge than Wikidata. OpenCyc is released under a free license and available to the public, but unlike Wikidata, OpenCyc is not supposed to be editable by the public. Freebase, acquired in 2010 by Google, is an online platform that allows communities to manage structured data [7]. Objects in Freebase are classified by *types* that prescribe what kind of data the object can have. For example, Freebase classifies Einstein as a *musical artist* since it would otherwise not be possible to refer to records of his speeches. Wikidata supports the use of arbitrary properties on all objects. Other differences to Wikidata are related to multi-language support, source information, and to the proprietary software used to run the site. The latter is critical for Wikipedia, which is committed to run on a fully open source software stack to allow anyone to fork the project.

Other approaches have aimed at extracting data from Wikipedia, most notably DBPedia [6] and Yago [13]. Both projects extract information from Wikipedia categories, and from the tabular *infoboxes* in the upper right of many Wikipedia articles. Additional mechanisms help to improve the extraction quality. Yago includes some temporal and spatial context information, but

neither DBpedia nor Yago extract source information.

Wikipedia data, obtained from the above projects or by custom extraction methods, has been used successfully to improve object search in Google's Knowledge Graph (based on Freebase) and Facebook's Open Graph, and in answering engines such as Wolfram Alpha [24], Evi [21], and IBM's Watson [10]. Wikipedia's geo-tags are also used by Google Maps. All of these applications would benefit from up-to-date, machine-readable data exports (e.g., Google Maps currently show India's Chennai district in the polar Kara Sea, next to Ushakov Island). Among the above applications, Freebase and Evi are the only ones that also allow users to edit or at least extend the data.

A Short History of Wikidata

Wikidata was launched October 2012. Editors could only create items and connect them to Wikipedia articles. In January 2013, three Wikipedias—first Hungarian, then Hebrew and Italian—started to connect to Wikidata. Meanwhile, the community had already created more than three million items. In February, the English Wikipedia followed, and in March all Wikipedias were connected to Wikidata.

Wikidata has received input from over 40,000 contributors so far. Since May 2013, Wikidata continuously had over 3,500 active contributors, i.e., contributors who make at least five edits within a month. These numbers make it one of the most active Wikimedia projects.

In March 2013, Lua was introduced as a scripting language to Wikipedia, which can be used to automatically create and enrich parts of articles, such as the infoboxes mentioned before. Lua scripts can access Wikidata, allowing Wikipedia editors to retrieve, process, and display data.

Many further features have been introduced in the course of 2013, and development is planned to continue in the foreseeable future.

Out of Many, One

The first challenge for Wikidata was to reconcile the 287 language editions of Wikipedia. For Wikidata to be truly multilingual, the object that represents *Rome* must be one and the same across all languages. Fortunately, Wikipedia already

Population	2,761,477		[edit]
	as of	2010	
	method	estimation	
	▸ 1 source		
			[add]

Figure 1: Screenshot of a complex statement as displayed in Wikidata

has a closely related mechanism: *language links*, displayed on the left of each article, connect articles in different languages. These links were created from user-edited text entries at the bottom of every article, leading to a quadratic number of links: *each* of the 207 articles about Rome contained a list of 206 links to all other articles about Rome—a total of 42,642 lines of text. By the end of 2012, Wikipedias in 66 languages contained more text for language links than for actual article content.

It would clearly be better to store and manage language links in a single location, and this was Wikidata's first task. For every Wikipedia article, a page has been created on Wikidata where links to related Wikipedia articles in all languages are managed. Such pages on Wikidata are called *items*. Initially, only a limited amount of data could be stored for each item: a list of language links, a label, a list of aliases, and a one-line description. Labels, aliases, and descriptions can be specified individually for currently up to 358 languages.

The Wikidata community has created bots to move language links from Wikipedia to Wikidata, and more than 240 million links could be removed from Wikipedia. Today, most language links displayed on Wikipedia are served from Wikidata. It is still possible to add custom links in an article, which is needed in the rare cases where links are not bi-directional: some articles refer to more general articles in other languages, while Wikidata deliberately connects only pages that cover the same subject. By importing language links, Wikidata obtained a huge set of initial items that are 'grounded' in actual Wikipedia pages.

Simple Data: Properties and Values
For storing structured data beyond text labels and language links, Wikidata uses a simple data model. Data is basically described by using property-value pairs. For example, the item for Rome might have a property *population* with value *2,777,979*. Properties are objects in their own right that have Wikidata pages with labels, aliases, and descriptions. In contrast to items, however, these pages are not linked to Wikipedia articles.

On the other hand, property pages always specify a datatype that defines which type of values the property can have. *Population* is a number, *has father* relates to another Wikidata item, and *postal code* is a string. This information is important to provide adequate user interfaces and to ensure that inputs are valid. There are only a small number of datatypes, mainly quantity, item, string, date and time, geographic coordinates, and URL. In each case, data is international, although its display may be language-dependent (e.g., the number 1,003.5 is written '1.003,5' in German and '1 003.5' in French).

Not-So-Simple Data
Property-value pairs are too simple for many cases. For example, Wikipedia states that the population of Rome was 2,761,477 *as of 2010* based on *estimations* published by *Istat*. Figure 1 shows how this could be represented in Wikidata. Even when leaving source information aside, the information can hardly be expressed in property-value pairs. One could use a property *estimated population in 2010*, or create an item *Rome in 2010* to specify a value for its *estimated population*—either solution is clumsy and impractical. As suggested by Figure 1, we would like the data

to contain a property *as of* with value *2010*, and a property *method* with value *estimation*. These property-value pairs do not refer to Rome, but to the assertion that Rome has a population of 2,761,477. We thus arrive at a model where the property-value pairs assigned to items can have additional subordinate property-value pairs, which we call *qualifiers*.

Qualifiers can be used to state contextual information, such as the validity time of an assertion. They can also be used to encode ternary relations that elude the property-value model. For example, to state that Meryl Streep played Margaret Thatcher in *The Iron Lady*, one could add to the item of the movie a property *cast member* with value *Meryl Streep*, and an additional qualifier 'role=Margaret Thatcher'.

These examples illustrate why we have decided to adopt an extensible set of qualifiers instead of restricting ourselves to the most common qualifiers, e.g., for temporal information. Indeed, qualifiers in their current form are an almost direct representation of data found in Wikipedia infoboxes today. This solution resembles known approaches of representing context information [18, 11]. It should not be misunderstood as a workaround to represent relations of higher arity in graph-based data models, since Wikidata statements do not have a fixed (or even bounded) arity in this sense [20].

Finally, Wikidata also allows for two special types of statements. First, it is possible to specify that the value of a property is unknown. For example, one can say that Ambrose Bierce's day of death is unknown rather than not saying anything about it. This clarifies that he is certainly
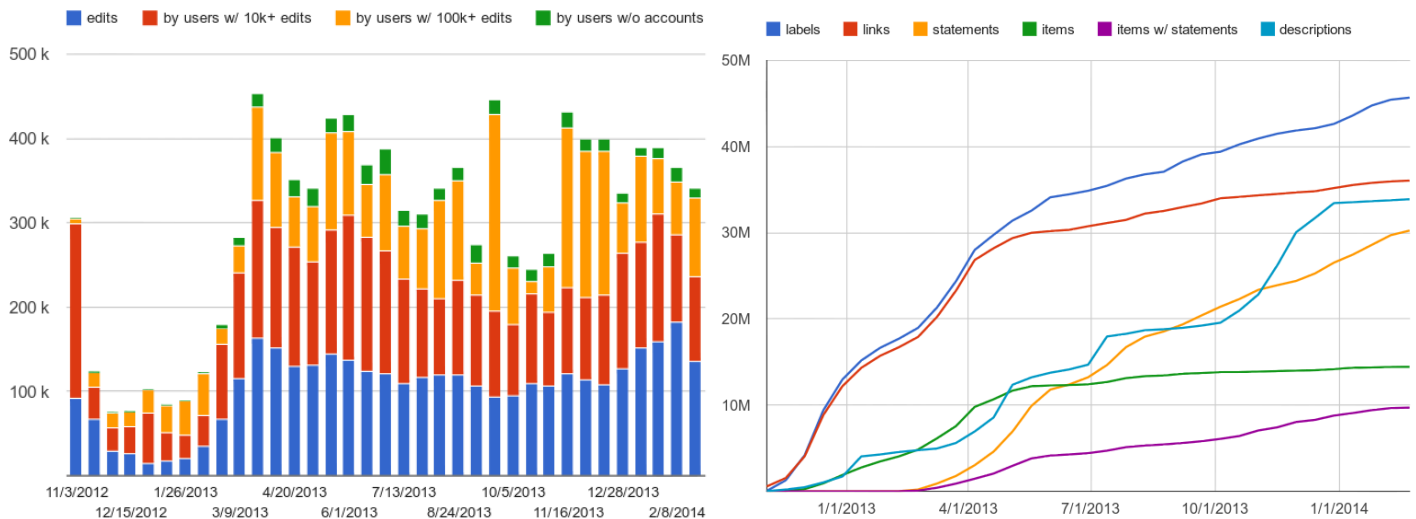
Figure 2: Growth of Wikidata: bi-weekly number of edits for different editor groups (left) and size of knowledge base (right)

not among the living. As the second additional feature, one can say that a property has no value at all, for example to state that Angela Merkel has no children. It is important to distinguish this situation from the common case that information is simply incomplete. It would be wrong to consider these two cases as special values. This becomes clear when considering queries that ask for items sharing the same value for a property—otherwise, one would have to conclude that Merkel and Benedict XVI have a common child.

The full data model and its expression in OWL/RDF can be found online [9].

Citation Needed

Property assertions, possibly with qualifiers, provide a rich structure to express arbitrary claims. In Wikidata, every such claim has a list of references to sources that support the claim. This agrees with Wikipedia's goal of being a secondary (or tertiary) source, that does not publish its own research but gathers information published in other primary (or secondary) sources.

There are many ways to specify a reference, depending on whether it is a book, a curated database, a website, or something entirely different. Moreover, some possible sources are represented by Wikidata items while others are not. Because of that, a reference is simply a list of property-value pairs, leaving the details of reference modeling to the community. Note that Wikidata does not automatically record

provenance [19], but rather provides for the structural representation of references.

Sources are also important as context information. Different sources often make contradicting claims, yet Wikidata should represent all views rather than choosing one 'true' claim. Combined with the context information provided by qualifiers (e.g., for temporal context), a large number of statements might be stored about a single property, such as population. To help manage this plurality, Wikidata allows contributors to optionally mark statements as *preferred* (for the most relevant, current statements) or *deprecated* (for irrelevant or unverified statements). Deprecated statements can be useful to Wikidata editors, to record erroneous claims of certain sources, or to keep statements that still need to be improved or verified. Like all content of Wikidata, these classifications are subject to community-governed editorial processes, similar to those of Wikipedia [1].

Wikidata in Numbers

Wikidata has grown significantly since its launch in October 2012. Some key facts about its current content are shown in Table 1. It has also become the most edited Wikimedia project, sporting 150–500 edits per minute, or half a million per day— about three times as many as the English Wikipedia. About 90% of these edits are made by bots that contributors have created for automating tasks, yet almost one million edits per month are made by humans. The left of Figure 2 shows the num-

ber of human edits during 14-day intervals. We highlight contributions of power users with more than ten or hundred thousand edits, respectively, as of February 2014; they account for most of the variation. The increase in March 2013 marks the official announcement of the site.

The right of Figure 2 shows the growth of Wikidata from its launch until February 2014. There are about 14.5 million items and 36 million language links. Essentially every Wikipedia article is connected to a Wikidata item today, so these numbers grow only slowly. In contrast, the number of labels, currently 45.6 million, continues to grow: there are more labels than Wikipedia articles. Almost 10 million items have statements, and more than 30 million statements have been created, using over 900 different properties. As expected, property usage is skewed: the most frequent property is *instance of* (P31, 5.6 million uses), which is used to classify items; one of the least frequent properties is P485 (133 uses), which connects a topic (e.g., Johann Sebastian Bach) with the institution that archives the topic (e.g., the Bach-Archiv in Leipzig).

The Web of Data

One of the promising developments in Wikidata is the community's reuse and integration of external identifiers from existing databases and authority controls, such as ISNI (International Standard Name Identifier), CALIS (China Academic Library & Information System), IATA (airlines and

Table 1. Some basic statistics about Wikidata as of February 2014

Supported languages	358	Statements	30,263,656	Edits	108,027,725
Labels	45,693,894	Statements with source	19,770,547	*Usage of datatypes:*	
Descriptions	33,904,616	Properties	920	– Wikidata items	20,135,245
Aliases	8,711,475	*Most-used properties:*		– Strings	7,589,740
Items	14,449,300	– instance of	5,612,339	– Geocoordinates	1,154,703
Items with statements	9,714,877	– country	2,018,736	– Points in time	912,287
Items with ≥5 statements	1,835,865	– taxon name	1,689,377	– Media files	386,357
Item with most statements:		Registered contributors	42,065	– URLs	75,614
– Rio Grande do Sul	511	with 5+ edits in Jan 2014	5,008	– Numbers *(new in 2014)*	9,842

airports), MusicBrainz (albums and performers), or HURDAT (North Atlantic hurricanes). These external IDs allow applications to integrate Wikidata with data from other sources, which remains under the control of the original publisher.

Wikidata is not the first project to reconcile identifiers and authority files from different sources. Other examples include VIAF for the bibliographic domain [3], GeoNames for the geographical domain [22], or Freebase [7]. Wikidata is linked to many of these projects, yet it also differs in terms of scope, scale, editorial processes, and author community.

The collected data is exposed in various ways.[1] Current per-item exports are available in JSON, XML, RDF, and several other formats. Full database dumps are created at intervals and supplemented by daily diffs. All data is licensed under CC0, putting the data into the public domain.

Every Wikidata entity is identified by a unique URI, such as http://www.wikidata.org/entity/Q42 for item Q42 (Douglas Adams). By resolving this URI, tools can obtain item data in the requested format (through content negotiation). This follows Linked Data standards for data publication [5], making Wikidata part of the Semantic Web [4] and supporting the integration of other Semantic Web data sources with Wikidata.

Wikidata Applications

The data in Wikidata lends itself to manifold applications on very different levels.

Language Labels and Descriptions. Wikidata provides labels and descriptions for many terms in different languages. These can be used to present information to international audiences. In contrast to common dictionaries, Wikidata covers a large number of named entities, such as names for places, chemicals, plants, and specialist terms, which can be very difficult to translate. Many data-centric views can be translated trivially term by term—think of maps, shopping lists, or ingredients of dishes on a menu—assuming that all items are associated with suitable Wikidata IDs.

Identifier Reuse. Item IDs can be used as language-independent identifiers to facilitate data exchange and integration across application boundaries. By referring to Wikidata items, applications can provide unambiguous definitions for the terms they use, which at the same time are the entry point to a wealth of related information. Wikidata IDs thus resemble Digital Object Identifiers (DOIs), but emphasizing (meta)data beyond online document locations, and using another social infrastructure for ID assignment. Wikidata IDs are stable: IDs do not depend on language labels, items can be deleted but IDs are never reused, and the links to other datasets and sites further increase stability. Besides providing a large collection of IDs, Wikidata also provides means to support contributors in selecting the right ID by displaying labels and descriptions—external applications can use the same functionality through the same API.

Accessing Wikidata. The information collected by Wikidata is interesting in its own right, and many applications can be built to access this information more conveniently and effectively. Applications created so far include generic data browsers like the one shown in Figure 3, and special-purpose tools including two genealogy viewers, a tree of life, a table of elements, and various mapping tools.[2] Applications can use the Wikidata API to browse, query, and even edit data. If simple queries are not enough, a dedicated copy of (parts of) the data is needed; it can be obtained from regular dumps and possibly be updated in real-time by following edits on Wikidata.

Enriching Applications. Many applications can be enriched by embedding information from Wikidata directly into their interfaces. For example, a music player might want to fetch the portrait of the artist just being played. In contrast to earlier uses of Wikipedia data, e.g., in Google Maps, it is unnecessary to extract and maintain the data. Such lightweight data access is particularly attractive for mobile apps. In other cases, it is useful to preprocess data to integrate it into an application. For example, it would be easy to extract a file of all German cities together with region and post code range, which could then be used in any application. Such derived data can be used and redistributed online or in software, under any license, even in commercial contexts.

Advanced Analytics. Information in Wikidata can further be analyzed to derive new insights beyond what is already stated. An important approach in this area is logical reasoning, where information about general relationships is used to derive additional facts. For example, Wikidata's property *grandparent* is obsolete since its value can be inferred from values of properties *father* and *mother*. If we are generally interested in ancestors, then a transitive closure needs to be computed. This is relevant for many hierarchical, spatial, and partonomical relations. Other types of advanced analytics include statistical evaluations, both of the data and of the incidental metadata collected in the system.

---

[1] See http://www.wikidata.org/wiki/Wikidata:Data_access
[2] An incomplete list is at http://www.wikidata.org/wiki/Wikidata:Tools

Figure 3: Wikidata in external applications: the data browser 'Reasonator' (http://tools.wmflabs.org/reasonator/)

For example, one can readily analyze article coverage by language [12], or the gender balance of persons with Wikipedia articles [14]. Like Wikipedia, Wikidata provides plenty of material for researchers to study.

These are only the most obvious approaches of exploiting the data, and many unforeseen uses can be expected. Wikidata is still very young and the data is far from complete. We look forward to new and innovative applications made possible by Wikidata and its development as a knowledge base [23].

Future Prospects

Wikidata is only at its beginning, with some crucial features still missing. These include support for complex queries, which is currently under development.

However, to predict the future of Wikidata, the plans of the development team might be less important than one would expect: the biggest open questions are about the evolution and interplay of the many Wikimedia communities. Will Wikidata earn the trust of the Wikipedia communities? How will the fact that such different Wikipedia communities, with their different languages and cultures, access, share,

and co-evolve the same knowledge base imprint on the way Wikidata is structured? How will Wikidata respond to the demands of communities beyond Wikipedia?

The influence of the community even extends to the technical development of the website and the underlying software. Wikidata is based on an open development process that invites contributions, and the site itself provides many extension points for user-created add-ons. Various interface features, e.g., for image embedding and multi-language editing, were designed and developed by the community. The community also developed ways to enrich the semantics of properties by encoding (soft) constraints such as 'items should not have more than one birthplace'. External tools gather this information, analyze the dataset for constraint violations, and publish the list of violations on Wikidata to allow editors to check if they are valid exceptions or errors.

These examples illustrate the close relationships between technical infrastructure, editorial processes, and content, and the pivotal role the community plays in shaping these aspects. The community, however, is as dynamic as Wikidata itself, based not on status or membership, but

on the common goal of turning Wikidata into the most accurate, useful, and informative resource possible. This goal provides stability and continuity, in spite of the fast-paced development, while allowing anyone interested to take part in defining the future of Wikidata.

Wikipedia is one of the most important websites today: a legacy that Wikidata still has to live up to. Within a year, Wikidata has already become an important platform for integrating information from many sources. In addition to this primary data, Wikidata also aggregates large amounts of incidental metadata about its own evolution and impact on Wikipedia. Wikidata thus has the potential to become a major resource for both research and the development of new and improved applications. Wikidata, the free knowledge base that everyone can edit, may thus bring us one step closer to a world in which everybody can freely share in the sum of all knowledge.

Acknowledgements

References

[1] Phoebe Ayers, Charles Matthews, and Ben Yates. *How Wikipedia works: And how you can be a part of it.* No Starch Press, 2008.

[2] Daniel J. Barrett. *MediaWiki*. O'Reilly Media, Inc., 2008.

[3] Rick Bennett, Christina Hengel-Dittrich, Edward T. O'Neill, and Barbara B. Tillett. VIAF (Virtual International Authority File): Linking Die Deutsche Bibliothek and Library of Congress name authority files. In *Proc. World Library and Information Congress: 72nd IFLA General Conference and Council.* IFLA, 2006.

[4] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, pages 96–101, May 2001.

[5] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):1–22, 2009.

[6] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia – A crystallization point for the Web of Data. *J. of Web Semantics*, 7(3):154–165, 2009.

[7] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proc. 2008 ACM SIGMOD Int. Conf. on Management of Data*, pages 1247–1250. ACM, 2008.

[8] Peter Buneman, James Cheney, Wang-Chiew Tan, and Stijn Vansummeren. Curated databases. In Maurizio Lenzerini and Domenico Lembo, editors, *Proc. 27th Symposium on Principles of Database Systems (PODS'09)*, pages 1–12. ACM, 2008.

[9] Wikimedia community. Wikidata: Data model. *Wikimedia Meta-Wiki*, 2012. https://meta.wikimedia.org/wiki/Wikidata/Data_model.

[10] David A. Ferrucci, Eric W. Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John M. Prager, Nico Schlaefer, and Christopher A. Welty. Building Watson: an overview of the DeepQA project. *AI Magazine*, 31(3):59–79, 2010.

[11] Ramanathan V. Guha, Rob McCool, and Richard Fikes. Contexts for the Semantic Web. In Sheila A. McIlraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *Proc. 3rd Int. Semantic Web Conf. (ISWC'04)*, volume 3298 of *LNCS*, pages 32–46. Springer, 2004.

[12] Scott A. Hale. Multilinguals and Wikipedia editing. arXiv:1312.0976 [cs.CY], 2013. http://arxiv.org/abs/1312.0976.

[13] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artif. Intell., Special Issue on Artificial Intelligence, Wikipedia and Semi-Structured Resources*, 194:28–61, 2013.

[14] Maximilian Klein and Alex Kyrios. VIAFbot and the integration of library data on Wikipedia. *code{4}lib Journal*, 2013. http://journal.code4lib.org/articles/8964.

[15] Markus Krötzsch, Denny Vrandečić, Max Völkel, Heiko Haller, and Rudi Studer. Semantic Wikipedia. *J. of Web Semantics*, 5(4):251–261, 2007.

[16] Douglas B. Lenat and Ramanathan V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, 1989.

[17] Bo Leuf and Ward Cunningham. *The Wiki way: quick collaboration on the Web*. Addison-Wesley Professional, 2001.

[18] Robert M. MacGregor. Representing reified relations in Loom. *J. Exp. Theor. Artif. Intell.*, 5(2-3):179–183, 1993.

[19] Luc Moreau. The foundations for provenance on the Web. *Foundations and Trends in Web Science*, 2(2–3):99–241, 2010.

[20] Natasha Noy and Alan Rector, editors. *Defining N-ary Relations on the Semantic Web*. W3C Working Group Note, 12 April 2006. Available at http://www.w3.org/TR/swbp-n-aryRelations/.

[21] William Tunstall-Pedoe. True Knowledge: open-domain question answering using structured knowledge and inference. *AI Magazine*, 31(3):80–92, 2010.

[22] Unxos GmbH. *GeoNames*, launched 2005. http://www.geonames.org, accessed Dec 2013.

[23] Denny Vrandečić. The Rise of Wikidata. *IEEE Intelligent Systems*, 28(4):90–95, 2013.

[24] Wolfram research. *Wolfram Alpha*, launched 2009. https://www.wolframalpha.com, accessed Dec 2013.

Denny Vrandečić (vrandecic@google.com) works at Google. He was the project director of Wikidata at Wikimedia Deutschland until September 2013.

Markus Krötzsch (markus.kroetzsch@tu-dresden.de) is lead of the Wikidata data model specification, and research group leader at TU Dresden.