Wikidata: The Making Of

Denny Vrandečić
Wikimedia Foundation
San Francisco, California, USA
Q18618629
denny@wikimedia.org

Lydia Pintscher
Wikimedia Deutschland
Berlin, Germany
Q18016466
lydia.pintscher@wikimedia.de

Markus Krötzsch
TU Dresden
Dresden, Germany
Q18618630
markus.kroetzsch@tu-dresden.de

ABSTRACT

Wikidata, now a decade old, is the largest public knowledge graph, with data on more than 100 million concepts contributed by over 560,000 editors. It is widely used in applications and research. At its launch in late 2012, however, it was little more than a hopeful new Wikimedia project, with no content, almost no community, and a severely restricted platform. Seven years earlier still, in 2005, it was merely a rough idea of a few PhD students, a conceptual nucleus that had yet to pick up many important influences from others to turn into what is now called Wikidata. In this paper, we try to recount this remarkable journey, and we review what has been accomplished, what has been given up on, and what is yet left to do for the future.

CCS CONCEPTS

• **Human-centered computing → Wikis**; • **Social and professional topics → Socio-technical systems**; **History of software**; • **Information systems → Wikis**.

KEYWORDS

Wikidata, knowledge graph, Wikibase, MediaWiki

1 INTRODUCTION

For many practitioners and researchers, Wikidata [68] simply is the largest freely available knowledge graph today. Indeed, with more than 1.4 billion statements about over 100 million concepts across all domains of human knowledge,[1] it is a valuable resource in many applications. Wikidata content is behind answers of smart assistants such as Alexa or Siri, is used in software and mobile apps (see Fig. 1), and enables research, e.g., in life sciences [38, 73], humanities and social sciences [33, 66, 76], artificial intelligence [1, 10, 49, 53, 57], and beyond [3, 46, 51].

[1]All statistics reported are current at the time of this writing. Up-to-date numbers are found at https://www.wikidata.org/wiki/Wikidata:Statistics.

However, Wikidata is much more than a data resource. It is, first and foremost, an international community of volunteers who subscribe to the goal of making free knowledge available to the world. It shares this and other goals with the wider *Wikimedia Movement*[2] to which Wikidata belongs. Indeed, Wikidata is also a project (and website) of the *Wikimedia Foundation*, along with sister projects such as Wikipedia and Wikimedia Commons, backed by dedicated staff to create and maintain the infrastructure that enables the work of the community.
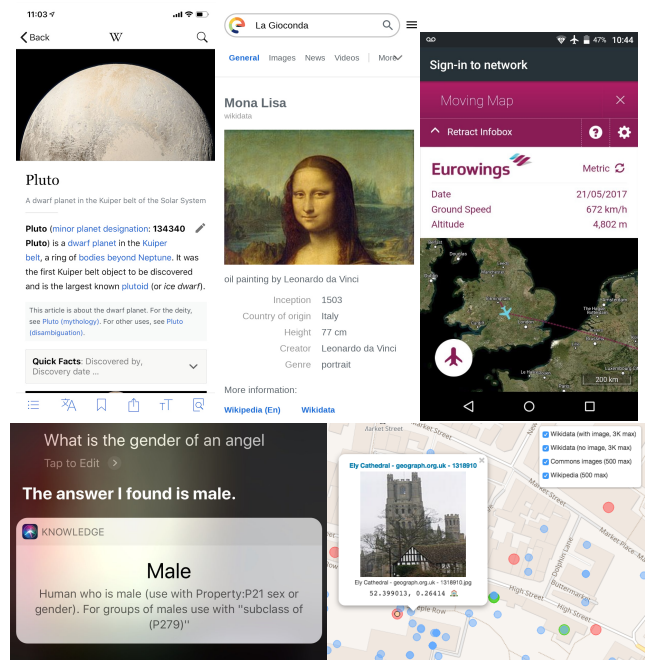

Figure 1: Apps using Wikidata (from upper left): Wikipedia iOS app, mobile search on e/OS/, in-flight app by Eurowings/Lufthansa Systems, Siri (historical glitch exposing Wikidata IDs), and WikiShootMe tool for Wikipedia editors

The complexity and scale of the endeavor may suggest that Wikidata was the result of a long and carefully prepared strategic plan of the Wikimedia Foundation, possibly in response to demands from the Wikipedia community. There is certainly some truth to that. However, the real history of how Wikidata was conceived, and how it eventually developed into its present form is not that straightforward: it involves a group of PhD students (naïve but optimistic[3]) a free software project that brought structured data

[2]https://meta.wikimedia.org/wiki/Wikimedia_movement
[3]We maintain that these are different qualities.

to thousands of wiki sites (successful, but not used for Wikidata), numerous funding proposals (some failed, some successful), and ideas from many different people, both in the Wikimedia movement and in the (Semantic) Web community.

Each step in this journey has also been witnessed by some or all of the authors of this paper, but a full account of these steps, their causes, and influences, has never been given in a coherent form. In this paper, we therefore embark upon the risky endeavor of recounting a history that is, in part, also our own. The result is nevertheless more than a piece of self-recorded *oral history*, since available online sources allow us to reconstruct not just what happened, but often also what the original plans and motivations have been, and how they have changed over time. Our subjective perspective will still play a major role in filling the gaps, offering explanations, and deriving objectives for the future.

Overall, we hope that our work can provide relevant insights not just about Wikidata, but also about the history of three influential ideals that have found their expression in many social, political, and technological developments of our time, especially on the Web:

(1) Community: the confidence that sensible people will work together to make the world a better place
(2) Sharing: the goal of making knowledge, and digital resources in general, freely available to every human
(3) Explication: the goal to formally specify information in explicit, unambiguous, and machine-processable ways

These ideals are neither universally accepted nor free of internal conflicts, but they continue to inspire. All three of them are closely tied to the development of the Web [8], to which they have also made important individual contributions: community is the basis of the *wiki* principle [34], sharing is the driving force of the open source and open knowledge movements, and explication through formal specification has motivated strong Web standards and the Semantic Web activity [7, 61]. Wikipedia naturally combines community and sharing, but Wikidata has pioneered the reconciliation of all three ideals.[4]

2 WIKIDATA AT TEN YEARS OF AGE

Before discussing its development further, we take a closer look at what Wikidata is today, to set it apart both from other activities and from its own former visions. As stated above, Wikidata is a knowledge graph, a community, an online platform, and a Wikimedia project. The aforementioned ideals are strongly represented in its design:

(1) Community: all content (data and schema) is directly controlled by an open community, not by the development team at Wikimedia Deutschland
(2) Sharing: the data is licensed under Creative Commons CC-0, which imposes no restrictions on usage or distribution
(3) Explication: content is structured according to its own data model, is exported in the RDF standard, and is open to machine reading/writing through APIs

In addition to these fundamentals, Wikidata is also characterized by several further design choices:

---

[4]Bernstein argues that community and formalization derive from partly opposing societal viewpoints that are typical of Northern America and Europe, respectively [8]. Within this view, Wikidata is a synthesis of American and European values.

Table 1: Statistics about Wikidata as of February 2023

Contributors	registered	565,000
	unregistered (distinct IPs)	1.6 million
	active per month	> 46,000
Items	total	>101 million
Properties	total	>10,800
	for external identifiers	> 7,800
Statements	total	>1.44 billion
	for external identifiers	>206 million
	per item (average)	14.3
Edits	overall	>1.8 billion
	per day (12 month avg)	699,000
Monthly page views (12 month avg)		420 million
Wikipedia articles using Wikidata		74%

(4) Multi-linguality: One Wikidata serves all languages; user-visible labels are translated, but underlying concepts and structures are shared; language-independent IDs are used
(5) Verifiability, not truth: Wikidata relies on external sources for confirmation; statements can come with references; conflicting or debated standpoints may co-exist
(6) Integration with Wikimedia: Wikidata is a data backbone for other Wikimedia projects (linking articles on the same topic across languages, providing data displayed in Wikipedia articles, supplying image tags for Wikimedia Commons, etc.)
(7) Identity provider: Wikidata concepts have stable, language-independent identifiers, linked with other resources (catalogs, archives, social networks, etc.) via external identifiers

These design choices distinguish Wikidata from many other structured knowledge collection efforts. Various projects rely on information extraction, partly from Wikipedia pages, most notably Yago [64], DBpedia [4], and Knowledge Vault [14]. Differences include the lack of direct community control, mono-linguality, and lack of verifiability (no references). Stronger similarities exist with the late Freebase [9], Metaweb's (and later Google's) discontinued knowledge graph community, and indeed some of this data was incorporated into Wikidata after the closing of that project [50]. Another related project is Semantic MediaWiki [26], on which we will have more to say later.

The data collected in most of these projects can also be considered knowledge graphs, i.e., structured data collections that encode meaningful information in terms of (typed, directed) connections between concepts. Nevertheless, the actual data sets are completely different, both in their vocabulary and their underlying data model. In comparison to other approaches, Wikidata has one of the richest graph formats, where each statement (edge in the graph) can have user-defined annotations (e.g., validity time) and references.

Today, Wikidata is at the core of the Wikimedia projects, a central resource of the world-wide knowledge ecosystem, and an integral part of technologies the world uses every day. Basic statistics are summarized in Table 1. Users of Wikidata's data include technology organizations (e.g., Google, IBM [19, 44], Quora [74], reddit, Wolfram Alpha [60], Apple, Amazon, OpenAI [53], Twitter [23]) and cultural and educational institutions (e.g., the Met [35], Smithsonian [62], Internet Archive [47], The Science Museum [15], dblp [55].

Wikidata's data is also used by Open Source and Open Culture projects (e.g., Wikipedia, MusicBrainz [39], OpenStreetMap [45], OpenArtBrowser[5], KDE [25], Wikitrivia[6], Scribe [17]) as well as civil society projects (e.g., OCCRP [65], Peppercat [11], OpenSanctions [48], GovDirectory [2], DataStory [36], EveryPolitician [12]).

Wikidata's data is used for a variety of tasks, including accessing basic information about a concept, machine learning, data cleaning and reconciliation, data exploration and visualization, tagging and entity recognition as well as internationalization of content. In addition, Wikidata is a hub in the Linked Data Web and beyond by connecting to over 7,500 other websites, catalogs, and other databases.

3 SEMANTIC WIKIPEDIA

Wikidata launched as a public web site in October 2012, but its true beginnings are much earlier, in May 2005. During the seven years between inception and launch, the design of Wikidata went through significant conceptual changes. This evolution was, however, driven not so much by deliberate strategic planning, but rather by close interactions with many people and communities as part of continuous efforts of making Wikidata a reality.

The first idea of what was to become Wikidata was born in early May 2005. Google was already strong, Skype had revolutionized (still voice-only) Internet telephony, Facebook was not fully public, and Twitter did not exist yet. Wikipedia, launched in 2001, was still not widely known, but its phase of explosive growth had started.[7] These were also formative years for the Wikimedia Movement, and the first ever Wikimania conference was to be held in August 2005 in Frankfurt am Main, Germany.[8]

Just one train-hour away, in Karlsruhe, a group of young PhD students were taking note. Markus Krötzsch, Max Völkel, and Denny Vrandečić had each recently joined the research group of Rudi Studer at University of Karlsruhe (now KIT), a leading location of Semantic Web research. Fascinated by the Wikipedia concept, and being early contributors,[9] it was natural to ask how the Semantic Web ideas of explicit specification and machine-readable processing could make a contribution. Vrandečić proposed to annotate links on Wikipedia pages, inspired by the notion of *typed links*, a well-known concept in hypertext that was also endorsed by Berners-Lee [6].

The result was the early concept of *Semantic Wikipedia*, a proposal to use annotations in wikitext markup for embedding structured data into Wikipedia articles [28]. Such integration of text and data was a popular concept in the early Semantic Web, and Wikipedia was perceived here as a miniature Web within which to realize these ideas. However, tying data to texts also enshrines mono-linguality, restricts machine-writability (since all data must also appear in text), and hinders verifiability (since it is hard to link data and references). None of these issues were perceived as very
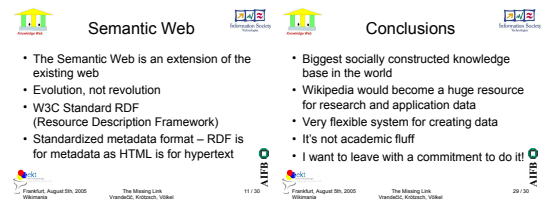

Figure 2: Slides from the presentation of Semantic Wikipedia at Wikimania 2005: pragmatic views and a call for action

problematic at the time,[10] whereas the seamless and gradual introduction of structured data management into existing workflows was considered essential.

Krötzsch and Vrandečić presented the idea at Wikimania on August 5th, 2005 (see Fig. 2). Certain of the convincing benefits of their vision, they called for volunteers to implement it – a task that Vrandečić, when asked, estimated to take about two weeks' effort. The German company *DocCheck* stepped up to donate this effort, leading to the first implementation of the software *Semantic MediaWiki* (SMW) [26, 29].

Looking back, the most striking aspect of this early history is how quickly the idea of a "Semantic Wikipedia" caught on and gathered support. In the 48 hours after their talk, Krötzsch and Vrandečić created a related community portal[11] with details on project goals, implementation plan, and envisioned applications (including "question answering based on Wikipedia (e.g. integrated in major web searching engines)".[12]) Within a month, the idea had gathered vocal supporters in the Web community, such as Tim Finin, Danny Ayers, and Mike Linksvayer.[13]. The SMW software saw its first release 0.1 on September 29th, 2005, with new mailing lists connecting to a growing user community. A first WWW paper was presented at WWW2006 in Edinburgh, Scotland [27].

This sudden success also reflects that Semantic Wikipedia resonated strongly with popular ideas of the time. Indeed, several *semantic wiki systems* (not related to Wikipedia) had been proposed around that time [13], and there was even a concurrent, completely independent (but not completely dissimilar) proposal for a "Semantic MediaWiki" by Hideaki Takeda and his research group, first published in October 2005 [41, 42]. The vision of a machine-readable Wikipedia also inspired researchers, which would later lead to DBpedia [4] and Yago [64] (both 2007).

Conversely, structured data had been gaining popularity within the Wikimedia Movement, e.g., in the German Wikipedia's "*Personendaten*" initiative.[14] In a peculiar historical coincidence, Erik Möller had recently proposed the idea of a Wikimedia project called *Wikidata*, conceived as a wiki-like database for several concrete application areas.[15] In the following years, Möller, Gerard Meijssen, and others pursued the *OmegaWiki* project (first named *WiktionaryZ*), which had an alternative approach to the data model [40,

---

[5]https://openartbrowser.org

[6]https://wikitrivia.tomjwatson.com

[7]Between Jan 2004 and Jan 2007, the number of editors with more than five monthly contributions to the English Wikipedia increased from 2,241 to over 100,000 (https://stats.wikimedia.org/).

[8]https://wikimania2005.wikimedia.org

[9]Krötzsch had edited Wikipedia since January 2003; Vrandečić since May 2003.

[10]It is curious that the Semantic Web vision of the early 2000s has generally favored machine-readability over machine-writability.

[11]See https://meta.wikimedia.org/wiki/Semantic_MediaWiki and its history

[12]https://meta.wikimedia.org/w/index.php?title=Semantic_MediaWiki/Envisaged_applications&oldid=188380

[13]https://simia.net/wiki/Semantic_Wikipedia

[14]https://de.wikipedia.org/wiki/Wikipedia:Personendaten

[15]See https://meta.wikimedia.org/wiki/Wikidata/Archive/Wikidata/historical and its page history.

43, 67], but was much more focused on multi-linguality from the beginning. It would still take years for these concepts to converge.

4 MOVING SIDEWAYS (2005–2010)

In the following years, the initial success of Semantic Wikipedia (the grand vision) gradually turned into a success of Semantic MediaWiki (the software). Fueled by the initiators' intensive development activities and community management, a growing user base was running their own Semantic MediaWiki sites in many application domains. The built-in query answering functionality turned out to be especially useful for many community projects outside of Wikipedia. Further developers joined, and the first *SMW User Meeting* in 2008 in Boston became the starting point of the regular *SMWCon* conference series, which is still ongoing today.[16]
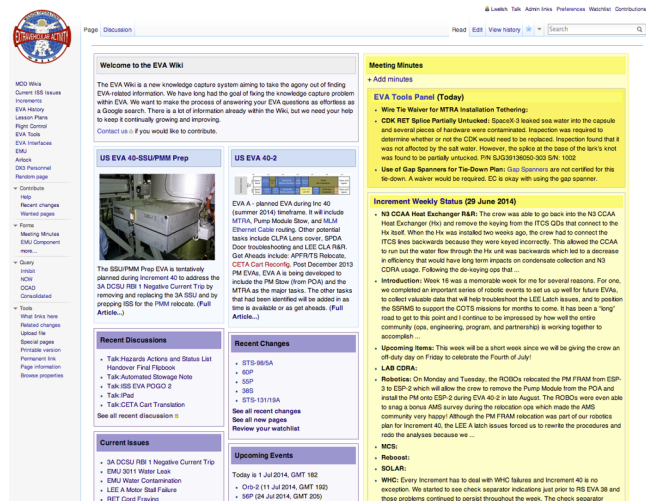

Figure 3: Screenshot of the internal EVA wiki, a Semantic MediaWiki instance used by astronauts at NASA

Practical experiences and user feedback from SMW also revealed aspects that the original concept of Semantic Wikipedia had overlooked or misjudged. Facilitated by new software extensions in the prospering SMW ecosystem, form-based input methods soon dominated over in-text annotations – a major deviation from the unity of data and text that was central to the Semantic Wikipedia concept. Moreover, it soon emerged that the provision of RDF-encoded data (fully conforming to Linked Data recommendations [5]) did not lead to relevant applications. Instead, the ability to embed query results into wiki pages was what motivated users to add structured data. Most development efforts between 2005 and 2010 were directed towards improving these *inline queries* in terms of power, performance, and presentation. Semantic wikis thus prospered, but, without inspiring data re-use beyond individual sites as expected for Wikipedia. SMW gradually diverged from its original goal.

Indeed, the praise that SMW won from researchers and practitioners had little impact on Wikipedia. SMW was presented at Wikimania each year (e.g. [30], often gathering significant audiences and positive feedback, yet the route into Wikipedia remained

unclear. For several years, editors and operators were occupied with running Wikipedia in a time of unprecedented growth. Potentially disruptive software changes were out of the question, and technical work focused on core functionality (UI, account management, discussion pages, backend performance). Even those more meager innovations were not always welcomed by the growing community, resulting in conflicts between the Wikimedia Foundation and contributors. Adding data management to Wikipedias' core tasks seemed a huge risk, especially as some communities became more conservative and less open to such changes. Even after many years, the only bits of structured data in Wikipedia came from a few uses of Microformats [24] – sparse records that would never form a knowledge graph.

SMW meanwhile was gathering interest elsewhere [26, 29, 31]. The large wiki host *Wikia* had made it a standard offer for its customers. Smaller IT companies offered services and extensions to turn it into a tool of corporate knowledge management. One of them was *Ontoprise*, a Karlsruhe-based SME subcontracted by Paul Allen's Vulcan Inc. under the leadership of Mark Greaves to adapt SMW for knowledge acquisition in the ambitious *Halo* project [18]. To support development and community outreach, Ontoprise hired a local computer science student, Lydia Pintscher.

5 EVOLUTION OF AN IDEA

While SMW was moving along its own trajectory, the greater goal was, however, not abandoned. Between 2005 and 2012, through interactions with many people, the original Semantic Wikipedia evolved to the first accurate concept of Wikidata.

Erik Möller, by then Deputy Director of the Wikimedia Foundation, was the driving force behind a major change: Vrandečić was still arguing to turn the individual Wikipedias semantic in 2009 (in particular to compare the graphs from the different language editions [58, 59, 71]), whereas Möller favored a single Wikidata for all languages. Already in Möller's original Wikidata proposal in 2004, he had envisioned a solution "to centrally store and manage data from all Wikimedia projects."[17] The resulting design combined this idea with the more fluid, graph-based data model of Semantic Wikipedia. Möller had also secured the domain for Wikidata, which was a major factor in eventually selecting this name.

Another important realization was that verifiability would have to play a central role. Vrandečić, Elena Simperl (then KIT), and Mathias Schindler (Wikimedia Deutschland) initiated research on the topic of knowledge diversity, which led up to the EU research project RENDER[18] (2010–2013). The project developed ideas for handling contradicting and incomplete knowledge, and analyzed Wikipedia to understand the necessity for such approaches [63].

Also in 2010, Krötzsch and Vrandečić had finished their Ph.D.s, with Krötzsch joining Ian Horrocks's group at the University of Oxford and Vrandečić following an invitation of Yolanda Gil to spend a 6-month sabbatical at ISI, University of Southern California. The first prototype for a verifiability-enabled semantic wiki platform, named *Shortipedia*, emerged from the collaboration of Vrandečić, Gil, Varun Ratnakar (ISI), and Krötzsch [70]. The prototype also

---

[16]https://www.semantic-mediawiki.org/wiki/SMWCon

[17]https://web.archive.org/web/20051223092827/http://meta.wikimedia.org/wiki/Wikidata

[18]http://render-project.eu

implemented language-independent identifiers with labels per language. Ideas were gradually converging towards Wikidata.

6 PROJECT PROPOSAL

Based on the long-standing interest in structured data around Wikimedia projects, Danese Cooper, then CTO of the Wikimedia Foundation, convened the Wikimedia Data Summit[19] in February 2011. Tim O'Reilly hosted the summit at the headquarters of O'Reilly in Sebastopol, CA. The invitation included representatives from the Wikimedia Foundation, Freebase (which had been acquired by Google the year prior), DBpedia, Semantic MediaWiki, R.V. Guha from Google, Mark Greaves from Paul Allen's Vulcan, and others. Many different ideas were discussed, but a rough consensus between some participants emerged, which would prompt Vrandečić to start writing a proposal for what at first was called *data.wikimedia.org*, but eventually would become *Wikidata*.

The project proposal draft[20] was presented to the community by Vrandečić at Wikimania 2011 in Haifa, Israel. At that event, Qamarniso Ismailova, an administrator of the Uzbek Wikipedia, and Vrandečić met. They married in August 2012. The Q prefix in QIDs, used as identifiers in Wikidata, is the first letter in her name.

Möller made it clear that the Wikimedia Foundation would be, at that point, reluctant to take on a project of this scale. Instead he identified the German chapter, Wikimedia Deutschland, as a good potential host for the development. Thanks to the on-going collaboration in RENDER, Pavel Richter, then Executive Director of Wikimedia Deutschland, took the proposal to WMDE's Board, which decided to accept Wikidata as a new Wikimedia project in June 2011, provided that sufficient funding would be available.[21] For Richter and Wikimedia Deutschland this was a major step, as the planned development team would significantly enlarge Wikimedia Deutschland, and necessitate a sudden transformation of the organization, which Richter managed in the years to come [56].

With the help of Lisa Seitz-Gruwell at the Wikimedia Foundation, they secured €1.3 Million in funding for the project: half from the Allen Institute for AI (AI2),[22] and a quarter each from Google and the Gordon and Betty Moore Foundation.[23] While looking for funding, at least one major donor dropped out because the project proposal insisted that the ontology of Wikidata had to be community-controlled, and would be neither pre-defined by professional ontologists nor imported from existing ontologies. Possible funders were also worried that the project did not plan to bulk-upload DBpedia to kick-start the content. Vrandečić was convinced that both of these requirements would not have had a positive effect on the organic growth of the community. Convinced the project would fail because of that, they dropped out.

7 EARLY DEVELOPMENT AND LAUNCH

The development of Wikidata began on April 1st, 2012 in Berlin. During the first months of development, the groundwork was laid

Figure 4: Initial Wikidata development team from April 2013; from left to right: John Blad, Abraham Taherivand, Tobias Gritschacher, Henning Snater, Jeroen De Dauw, Daniel Kinzler, Markus Krötzsch, Lydia Pintscher, Silke Weber, Denny Vrandečić, Daniel Werner, Katie Filbert, Jens Ohlig

to allow MediaWiki to store structured data. At the same time, discussions in the Wikimedia communities led to people laying with the Wikidata project their highest hopes, and biggest concerns, for what the project would mean for the future of Wikipedia and the larger Wikimedia Movement. There were fears and fantasies of complete automation of Wikipedia article writing, a forced uniformity and alignment across the different Wikipedia language editions, and the loss of nuance and cultural context in structured data. Fortuitously, Google announced the Knowledge Graph in May 2012, which had a lasting positive impact on the interest into Wikidata.

Wikidata launched on October 29, 2012. This initial launch was, intentionally, very limited. Users could create new identifiers for concepts (QIDs), label them in many languages, and link them to Wikipedia articles and other Wikimedia pages. Statements were not supported yet, and the collected links and labels were not used anywhere. The first community-created item was about Africa.

One surprisingly contentious aspect was the use of numeric QIDs. Numeric QIDs are still being questioned today, with proponents arguing that a qualified name such as dbp:Tokyo is easier to understand than Q1490. A major influence for preferring abstract[24] QIDs were discussions with Metaweb regarding their experience with Freebase. Moreover, most other online databases, authority files, and ontologies were also preferring abstract IDs. De-coupling a concept's name from its ID can increase stability (since IDs do not change if names do), but studies found that Wikipedia article titles often are rather stable identifiers [20]. More importantly, however, the founders of Wikidata did not want to use an anglo-centric solution, nor suggest the use of many different language-specific identifiers (such as dbp:東京都) for the same Item.

8 EARLY WIKIDATA (2013–2015)

Wikidata grew. Crucial functionality was added, the community grew, and the content was expanded alongside many initial data modeling discussions. This growth was intentionally managed to be slow and steady, in order to build a healthy project, created and supported by a sustainable community.

---

[19] https://meta.wikimedia.org/wiki/Data_summit_2011

[20] https://meta.wikimedia.org/wiki/Wikidata/Technical_proposal

[21] https://www.wikimedia.de/wp-content/uploads/2019/10/Beschlüsse-des-8.-Vorstandes.pdf

[22] Wikidata actually triggered the creation of AI2: Paul Allen's Vulcan Inc. could not legally provide charitable donations without a commercial contract, as required by Wikimedia, leading him to pursue the idea of an AI-related nonprofit that had been discussed for some time (Mark Greaves, personal communication).

[23] https://techcrunch.com/2012/03/30/wikipedias-next-big-thing-wikidata-a-machine-readable-user-editable-database-funded-by-google-paul-allen-and-others

[24] Some rare QIDs do have meaning: https://www.wikidata.org/wiki/Wikidata:Humour
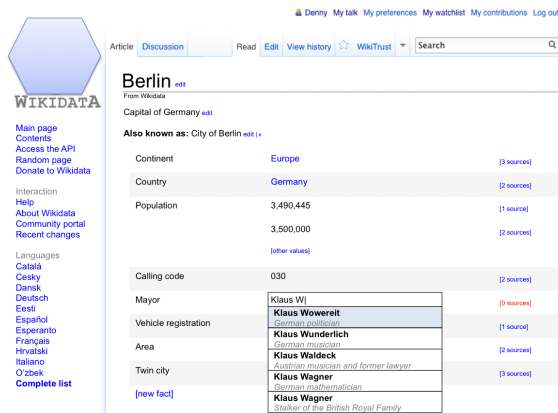
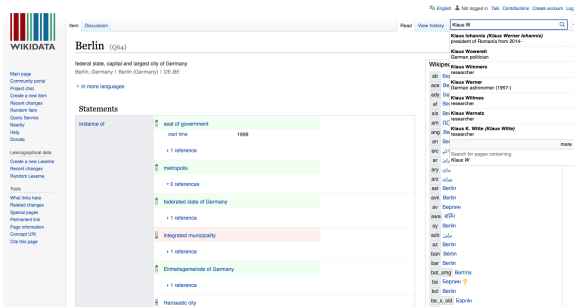Figure 5: Early mockup of the Wikidata UI

Figure 6: Screenshot of the Wikidata UI as of 2022

Among the important functionality added in the first months after launch were the support for wikis to use Wikidata's sitelinks rather than encode that locally (first Wikipedias – Hungarian, Hebrew and Italian – in January 2013, and English in February 2013). Wikipedias started deleting the sitelinks from their local articles. This led to a removal of more than 240 million lines of wikitext across Wikipedia language editions, which reduced the cross-wiki maintenance effort massively [72]. In some languages, these lines constituted more than half of the content of that Wikipedia language edition. In many languages, editing activity dropped dramatically at first, sometimes by 80%. But those edits were mostly from bots that were previously needed to synchronize links across languages. With those bots gone, humans were suddenly better able to 'see' each other and build a more meaningful community. In many languages, this eventually led to an increased community activity. In addition to reducing the maintenance burden on the Wikipedias, this also lead to the creation of Items on Wikidata for a large number of general concepts represented by each of these articles, and thereby helped bootstrap the content of Wikidata.

Further development introduced properties and basic statements (January 2013), as well as basic support for including data from Wikidata into Wikipedia articles (April 2013). The editor community started rallying around the tasks that could be done with the limited functionality and started forming task forces (later becoming WikiProjects) to collect and expand data around topics such as countries and Pokémon, or to improve the language coverage for
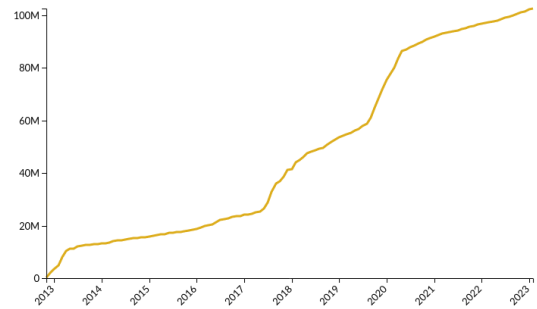

Figure 7: Number of pages on Wikidata: Wikidata reached 100 million entities just in the week before its tenth birthday
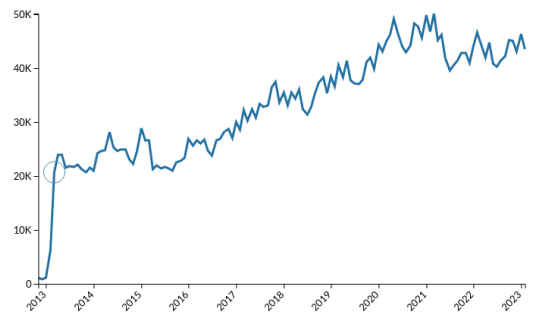

Figure 8: Monthly active contributors on Wikidata; the circle indicates when the initial import of sitelinks finished

certain languages.[25] This initial editor community was a healthy mix of people who were doing similar work on Wikipedia and found Wikidata to be a better fit for their type of work, some open data enthusiasts, and Semantic Web people who were excited by the idea of Wikimedia embracing (some of) their ideas and by what this would enable going forward.

Along the way, the skepticism of some (though not all) Wikimedians could be addressed by Wikidata showing its benefits and potential, and by the care that had been put into its foundational design. The centralization of sitelinks brought a lot of goodwill to Wikidata.[26] It helped that early on it was decided that Wikidata would not be forced upon any Wikimedia project, but that instead it would be up to the editor community on each wiki to decide where and how they would make use of data from Wikidata.

It has been a challenge to make the idea of a knowledge graph accessible and attractive to an audience that is not familiar with the ideas of the Semantic Web. Data is abstract, and it takes creativity and effort to see the potential in linking this data and making it machine-readable. A few key applications were instrumental in sparking excitement by showing what is and will become possible once Wikidata grew. Chief among the people who made this possible was Magnus Manske, who developed Reasonator,[27] an alternative view on Wikidata; Wiri,[28] an early question answering demo; and Wikidata Query, the first query tool for Wikidata.

---

[25]In January 2013 the labels and descriptions task force succeeded in creating English labels for all of the first 20,000 Items.
[26]Former Board of Trustee member Phoebe Ayers stated that *"the sitelinks alone were worth the price of admission"*.
[27]https://reasonator.toolforge.org
[28]https://magnus-toolserver.toolforge.org/thetalkpage

Figure 9: Wikidata won the Open Data Award in 2014; from left to right: Sir Nigel Shadbolt, Lydia Pintscher, Magnus Manske, Sir Tim Berners-Lee

In addition to inspiring people's imagination, it was also necessary to support the editors with specialized and large-scale editing tools to be able to create and maintain the vast knowledge graph. The development team focused on the core of Wikidata, and community members stepped up to the task of building these tools around the core. Here too, Manske was chief among them. He created tools such as Mix'n'match,[29] for matching Wikidata Items to entries in other catalogs; Terminator,[30] for gathering translations for Items in missing languages; The Wikidata Game,[31] for answering a question that will result in an edit on Wikidata[32]; and – maybe most importantly – QuickStatements,[33] that significantly lowered the bar for mass edits by non-technical editors.

During this time, it also became apparent that more support for the editors was needed to define 'rules' around the data without losing the flexibility and openness of the project. In July 2015, the Property constraint system was introduced, which enabled editors to specify in a machine-readable way how each of the thousands of Properties should be used.

In September 2015, the initial Wikidata Query tool by Manske had served its purpose as a feasibility study and demo, and the Wikidata Query Service (WDQS) was launched.[34] WDQS is a Blazegraph-based SPARQL endpoint that gives access to the RDF-ized version [16, 21] of the data in Wikidata in real-time, through live updates [37]. Its goal is to enable applications and services on top of Wikidata, as well as to support the editor community, especially in improving data quality. Originally, Vrandečić had not planned for a SPARQL query service, as he did not think that any of the available Open Source solutions would be performant enough to support Wikidata. Fortunately he was wrong, and today the SPARQL query service has become an integral part of the Wikidata ecosystem.[35] In particular, the query service allows for the creation of beautiful, even interactive visualizations directly from a query, such as maps, galleries, and graphs (see Figure 10). The service supports federation with other query endpoints, and allows for downloading the results in various formats. Reaching the 2020s, however, the query service
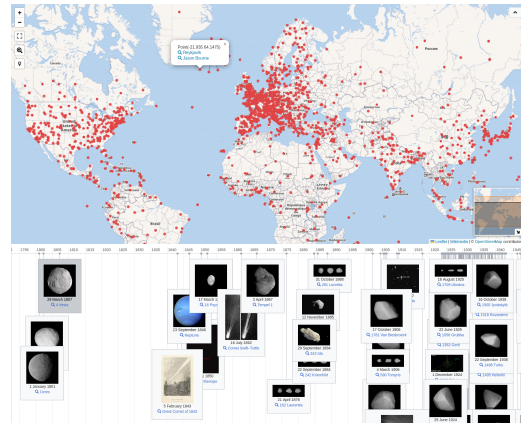

Figure 10: Example query visualizations from WDQS: locations of movie narratives (*above*), and a timeline of space discoveries (*below*)

has started to become a bottleneck, as the growth of Wikidata has outpaced the development of Open Source triplestores [75].

9 TEENAGE WIKIDATA (2015-2022)

In 2016, Google closed down Freebase and helped with the migration of the data to Wikidata [50]. The Wikidata community picked up the data carefully and slowly, and ensured that the influx of data would not push beyond their capacity to maintain it.

While Wikidata was always imagined to be useful outside of Wikipedia, its development had started out with the focus of providing a backbone for Wikipedia. This very soon expanded to the other Wikimedia projects, as well as data consumers outside of Wikimedia looking for general purpose data. But that is not the only expansion that Wikidata went through.

In early 2018, Wikidata was extended to also be able to cover lexicographical data, in order to provide a machine-readable dictionary. In late 2018, Wikimedia Commons was enhanced with the ability to record machine-readable data about its media files, based on Wikidata's concepts and technology.

The newest wave of expansion is the Wikibase Ecosystem, where groups and organizations outside Wikimedia use the underlying software that powers Wikidata (called Wikibase[36]) to run their own knowledge graphs, which are often highly inter-connected with Wikidata and other Wikibase instances, as well as other resources on the Linked Data Web.

10 OUTLOOK

The first ten years of Wikidata are just its beginning. There is hopefully much more to come, but also much more still to do.

Indeed, even the original concept has not been fully realized yet. The initial Wikidata proposal (Section 6) was split in three phases: first sitelinks, second statements, third queries. The third phase, though, has not yet been realized. It was planned to allow the community to define queries, to store and visualize the results in Wikidata, and to include these results in Wikipedia. This would have served as a forcing function to increase the uniformity of Wikidata's structure.

---

[29]https://mix-n-match.toolforge.org

[30]https://wikidata-terminator.toolforge.org

[31]https://www.wikidata.org/wiki/Wikidata:The_Game

[32]Example questions include "Is this Wikipedia article about a human?" or "Is this image a good representation of this concept?"

[33]https://quickstatements.toolforge.org

[34]https://lists.wikimedia.org/hyperkitty/list/wikidata@lists.wikimedia.org/thread/N2HPRCYIWGLM2IDTNCHQLNY574H5ZEQR/

[35]Malyshev et al. have curated a large, freely available dataset of anonymized WDQS queries and analyzed practical usage [37].

[36]https://wikiba.se

Figure 11: One of the cakes made for Wikidata's tenth birthday: the two QIDs refer to the song *Happy Birthday To You* and to Wikidata

Indeed, data uniformity and coherency has emerged as one of the big challenges that Wikidata has yet to address. By selecting a flexible, statement-centric data model – inspired by SMW, and in turn by RDF – Wikidata does not enforce a fixed schema upon groups of concepts. This is a maximal departure from the historic Wikidata plan (Section 3), and even from the more flexible (but still template-based) Freebase. There are advantages to such flexibility (e.g., Freebase struggled with evolving schemas or unexpected needs), but it also leads to reduced coherence and uniformity across groups of similar concepts, which is an obstacle to re-use.

On-going and future developments may help to address this, while unlocking additional new uses of the data. Two notable Wikimedia projects under current development are Wikifunctions and Abstract Wikipedia [69], both led by Vrandečić. These closely-related projects have a number of goals. Most prominently, Abstract Wikipedia is working towards extending knowledge representation beyond Wikidata such that one can abstractly capture the contents and structure of Wikipedia articles. Building on top of the data and lexicographic knowledge in Wikidata, these abstract representations will then be used to generate encyclopedic content in many more languages, providing a baseline of knowledge in the hundreds of languages of Wikipedia. This will also require significantly more lexicographic knowledge of Wikidata than currently available (about 1,000,000 Lexemes in 1,000 languages as of February 2023).

Wikifunctions in turn is envisioned as a wiki-based repository of executable functions, described in community-curated source code. These functions will in particular be used to access and transform data in Wikidata, in order to generate views on the data. These views – tables, graphs, text – can then be integrated into Wikipedia. This is a return to the goals of the original Phase 3, which would increase both the incentives to make the data more coherent, and the visibility and reach of the data as such. This may then lead to improved correctness and completeness of the data, since only data that is used is data that is good (a corollary to Linus's law of *"given enough eyeballs, all bugs are shallow"* [54]).

Returning to Wikidata itself, there are also many important tasks and developments still ahead. Editors need continued support to maintain data quality and increase coherence, an everlasting challenge in an open and dynamic system. Together with easier access methods, this should enable more applications, services, and research on top of the data and increase Wikidata's impact further.

In addition, Wikidata still has a long way to go to fully realize its potential as a support system for the Wikimedia projects.

Another aspect of Wikidata that we think needs further development is how to more effectively share semantics – within Wikidata itself, with other Wikimedia projects, and with the world in general. Wikidata is not based on a standard semantics such as OWL [22], although community modeling is strongly inspired by some of the expressive features developed for ontologies. The intended modeling of data is communicated through documentation on wikidata.org, shared SPARQL query patterns, and Entity Schemas in ShEx [52]. Nevertheless, the intention of modeling patterns and individual statements often remains informal, vague, and ambiguous. As Krötzsch argued in his ISWC 2022 keynote [32], a single, fixed semantic model could not be enough for all uses and perspectives required for Wikidata (or the Web as a whole), yet some sufficiently formal, unambiguous, and declarative way of sharing intended interpretations is still needed. A variety of powerful knowledge representation languages could be used for this purpose, but we still lack both infrastructure and best practices to use them effectively in such complex applications.

The above are mainly the wishes and predictions of the authors. The beauty of Wikidata is, however, that many people have used the system and data in ways we never have imagined, and we hope and expect that the future will continue to surprise us.

ACKNOWLEDGMENTS

REFERENCES

[1] Oshin Agarwal, Heming Ge, Siamak Shakeri, and Rami Al-Rfou. 2021. Knowledge Graph Based Synthetic Corpus Generation for Knowledge-Enhanced Language Model Pre-training. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tür, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, 3554–3565. https://doi.org/10.18653/v1/2021.naacl-main.278

[2] Jan Ainali. 2022. *Getting all the government agencies of the world structured in Wikidata*. Wikimedia. Retrieved 14 Nov 2022 from https://diff.wikimedia.org/2022/03/06/getting-all-the-government-agencies-of-the-world-structured-in-wikidata

[3] Stacy Allison-Cassin and Dan Scott. 2018. Wikidata: a platform for your library's linked open data. *The Code4Lib Journal* Issue 40 (may 2018). https://journal.code4lib.org/articles/13424

[4] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A nucleus for a web of open data. In *Proceedings of the 2007 ISWC International Semantic Web Conference*. Springer, 722–735.

[5] Tim Berners-Lee. 2006. Linked Data. https://www.w3.org/DesignIssues/LinkedData.html.

[6] Tim Berners-Lee and Mark Fischetti. 1999. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper, San Francisco, CA, USA.

[7] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American* 284, 5 (2001), 34–43.

[8] Mark Bernstein. 2022. On The Origins Of Hypertext In The Disasters Of The Short 20th Century. In *Proc. of the ACM Web Conference 2022 (WWW'22)*, Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini (Eds.). ACM, 3450–3457.

[9] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, 1247–1250.

[10] Jan A. Botha, Zifei Shan, and Daniel Gillick. 2020. Entity Linking in 100 Languages. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 7833–7845. https://doi.org/10.18653/v1/2020.emnlp-main.630

[11] Tony Bowden. 2022. *The CIA lost track of who runs the UK, so I picked up the slack*. OpenSanctions. Retrieved 14 Nov 2022 from https://www.opensanctions.org/articles/2022-01-18-peppercat

[12] Tony Bowden and Lucy Chambers. 2017. *EveryPolitician: the road ahead*. mySociety. Retrieved 14 Nov 2022 from https://www.mysociety.org/2017/07/05/everypolitician-the-road-ahead

[13] François Bry, Sebastian Schaffert, Denny Vrandečić, and Klara Weiand. 2012. Semantic wikis: Approaches, applications, and perspectives. In *Reasoning Web International Summer School*. Springer, 329–369.

[14] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*, Sofus A. Macskassy, Claudia Perlich, Jure Leskovec, Wei Wang, and Rayid Ghani (Eds.). ACM, 601–610.

[15] Kalyan Dutia and John Stack. 2021. Heritage connector: A machine learning framework for building linked open data from museum collections. *Applied AI Letters* 2, 2 (2021), e23.

[16] Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. 2014. Introducing Wikidata to the Linked Data Web. In *Proceedings of the 13th International Semantic Web Conference (ISWC'14) (LNCS, Vol. 8796)*, Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig A. Knoblock, Denny Vrandečić, Paul T. Groth, Natasha F. Noy, Krzysztof Janowicz, and Carole A. Goble (Eds.). Springer, 50–65.

[17] Elisabeth Giesemann. 2022. *Lexicographical Data for Language Learners: The Wikidata-based App Scribe*. Wikimedia Deutschland. Retrieved 14 Nov 2022 from https://tech-news.wikimedia.de/en/2022/03/18/lexicographical-data-for-language-learners-the-wikidata-based-app-scribe

[18] David Gunning, Vinay Chaudhri, Peter Clark, Ken Barker, Shaw Chaw, Mark Greaves, Benjamin Grosof, Alice Leung, David Mcdonald, Sunil Mishra, John Pacheco, Bruce Porter, Aaron Spaulding, Dan Tecuci, and Jing Tien. 2010. Project Halo Update - Progress Toward Digital Aristotle. *AI Magazine* 31, 3 (09 2010), 33–58. https://doi.org/10.1609/aimag.v31i3.2302

[19] Oktie Hassanzadeh. 2022. Building a Knowledge Graph of Events and Consequences Using Wikipedia and Wikidata. In *Proceedings of the Wiki Workshop at The Web Conference 2022*. ACM, 6 pages.

[20] Martin Hepp, Katharina Siorpaes, and Daniel Bachlechner. 2007. Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management. *IEEE Internet Computing* 11, 5 (2007), 54–65. https://doi.org/10.1109/MIC.2007.110

[21] Daniel Hernández, Aidan Hogan, and Markus Krötzsch. 2015. Reifying RDF: What Works Well With Wikidata?. In *Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2015) (CEUR Workshop Proceedings, Vol. 1457)*. CEUR-WS.org, 32–47.

[22] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph (Eds.). 27 October 2009. *OWL 2 Web Ontology Language: Primer*. W3C Recommendation, Boston, MA, USA. Available at http://www.w3.org/TR/owl2-primer/.

[23] Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. Algorithmic amplification of politics on Twitter. *Proc. Natl. Acad. Sci. USA* 119, 1 (2022), e2025334119.

[24] Rohit Khare and Tantek Çelik. 2006. Microformats: a pragmatic path to the semantic web. In *Proceedings of the 15th international conference on World Wide Web*. ACM, 865–866.

[25] Volker Krause. 2018. *KDE Itinerary - Static Knowledge*. KDE. Retrieved 14 Nov 2022 from https://www.volkerkrause.eu/2018/09/15/kde-itinerary-static-knowledge.html

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

[27] Markus Krötzsch, Max Völkel, Heiko Haller, Rudi Studer, and Denny Vrandečić. 2006. Semantic Wikipedia. In *Proceedings of the 15th International Conference on World Wide Web (WWW'06)*. ACM, 585–594.

[28] Markus Krötzsch, Max Völkel, and Denny Vrandečić. 2005. Wikipedia and the Semantic Web: The Missing Links. In *Proceedings of Wikimania 2005 - The 1st International Wikimedia Conference*. Wikimedia Foundation, 15 pages. https://meta.wikimedia.org/wiki/Wikimania05/Paper-MK2.

[29] Markus Krötzsch, Max Völkel, and Denny Vrandečić. 2006. Semantic MediaWiki. In *Proceedings of the 5th International Semantic Web Conference (ISWC'06) (LNCS, Vol. 4273)*, Isabel Cruz et al. (Ed.). Springer, 935–942.

[30] Markus Krötzsch, Max Völkel, and Denny Vrandečić. 2007. Wikipedia and the Semantic Web, Part II. In *Proceedings of the 2nd International Wikimedia Conference (Wikimania'06)*, Phoebe Ayers and Nicholas Boalch (Eds.). Wikimedia Foundation, 15 pages.

[31] Markus Krötzsch and Denny Vrandečić. 2008. Semantic Wikipedia. In *Social Semantic Web*. Springer, Berlin, Germany, 393–422.

[32] Markus Krötzsch. 2022. Data, Ontologies, Rules, and the Return of the Blank Node. https://www.youtube.com/watch?v=ryxusv0s604. Keynote at the 21st International Semantic Web Conference.

[33] Morgane Laouenan, Palaash Bhargava, Jean-Benoît Eyméoud, Olivier Gergaud, Guillaume Plique, and Etienne Wasmer. 2022. A cross-verified database of notable people, 3500BC-2018AD. *Scientific Data* 9, 1 (June 2022), 19 pages. https://doi.org/10.1038/s41597-022-01369-4

[34] Bo Leuf and Ward Cunningham. 2001. *The Wiki way: Quick collaboration on the Web*. Addison-Wesley, Boston, MA, USA.

---

[37]https://commons.wikimedia.org/wiki/File:Wikidata_Fotos_183.JPG
[38]https://www.flickr.com/photos/ukodi/15101674023/
[39]https://commons.wikimedia.org/wiki/File:Q167545_and_Q2013_on_cake.jpg

[35] Andrew Lih. 2019. *Combining AI and Human Judgment to Build Knowledge about Art on a Global Scale*. The Met. Retrieved 14 Nov 2022 from https://www.metmuseum.org/blogs/now-at-the-met/2019/wikipedia-art-and-ai

[36] Robin Linderborg. 2021. *How we're tracking elections in symbiosis with Wikidata*. Datastory. Retrieved 14 Nov 2022 from https://www.datastory.org/blog/tracking-the-worlds-elections

[37] Stanislav Malyshev, Markus Krötzsch, Larry González, Julius Gonsior, and Adrian Bielefeldt. 2018. Getting the most out of Wikidata: semantic technology usage in Wikipedia's knowledge graph. In *Proceedings of the International Semantic Web Conference 2018 (ISWC'18)*. Springer, 376–394.

[38] Magnus Manske, Ulrike Böhme, Christoph Püthe, and Matt Berriman. 2019. GeneDB and Wikidata. *Wellcome Open Research* 4, 114 (2019), 8 pages. https://doi.org/10.12688/wellcomeopenres.15355.2

[39] Ian McEwen. 2015. *Downstream Wikipedia link usage and migration to Wikidata*. MetaBrainz. Retrieved 14 Nov 2022 from https://blog.metabrainz.org/2015/05/08/downstream-wikipedia-link-usage-and-migration-to-wikidata

[40] Gerard Meijssen. 2009. The Philosophy behind OmegaWiki. *Lexicography at a crossroads: Dictionaries and encyclopedias today, lexicographical tools tomorrow* 90 (2009), 91.

[41] Hendry Muljadi and Hideaki Takeda. 2005. Semantic Wiki as an Integrated Content and Metadata Management System. Poster/Demo at the 4th International Semantic Web Conference (ISWC'05),http://www-kasm.nii.ac.jp/papers/takeda/05/hendry05iswc.pdf.

[42] Hendry Muljadi, Hideaki Takeda, Jiro Araki, Shoko Kawamoto, Satoshi Kobayashi, Yoko Mizuta, Sven Minoru Demiya, Satoshi Suzuki, Asanobu Kitamoto, Yasuyuki Shirai, et al. 2005. Semantic MediaWiki: A user-oriented system for integrated content and metadata management system. In *Proceedings of the IADIS International Conference WWW/Internet*. IADIS, 19–22.

[43] Erik Möller. 2006. *Die heimliche Medienrevolution: Wie Weblogs, Wikis und freie Software die Welt verändern* (2 ed.). Heise Medien, Hannover, Germany.

[44] Sumit Neelam, Udit Sharma, Hima Karanam, Shajith Ikbal, Pavan Kapanipathi, Ibrahim Abdelaziz, Nandana Mihindukulasooriya, Young-Suk Lee, Santosh K. Srivastava, Cezar Pendus, Saswati Dana, Dinesh Garg, Achille Fokoue, G. P. Shrivatsa Bhargav, Dinesh Khandelwal, Srinivas Ravishankar, Sairam Gurajada, Maria Chang, Rosario Uceda-Sosa, Salim Roukos, Alexander G. Gray, Guilherme Lima, Ryan Riegel, Francois P. S. Luus, and L. Venkata Subramaniam. 2022. A Benchmark for Generalizable and Interpretable Temporal Question Answering over Knowledge Bases. *CoRR* abs/2201.05793 (2022), 7 pages. arXiv:2201.05793 https://arxiv.org/abs/2201.05793

[45] Minh Nguyen. 2016. *Connecting OpenStreetMap and Wikidata*. Mapbox. Retrieved 14 Nov 2022 from https://blog.mapbox.com/connecting-openstreetmap-and-wikidata-232f0c412926

[46] Finn Årup Nielsen, Daniel Mietchen, and Egon Willighagen. 2017. Scholia, Scientometrics and Wikidata. In *The Semantic Web: ESWC 2017 Satellite Events*, Eva Blomqvist, Katja Hose, Heiko Paulheim, Agnieszka Ławrynowicz, Fabio Ciravegna, and Olaf Hartig (Eds.). Springer International Publishing, 237–259.

[47] Nick Norman. 2020. *Amplifying the Voices Behind Books With the Power of Data*. Internet Archive. Retrieved 14 Nov 2022 from https://blog.openlibrary.org/2020/09/02/amplifying-the-voices-behind-books

[48] OpenSanctions. 2022. *We're now integrating persons of interest from Wikidata!* OpenSanctions. Retrieved 14 Nov 2022 from https://www.opensanctions.org/articles/2022-01-25-wikidata

[49] Malte Ostendorff, Peter Bourgonje, Maria Berger, Julián Moreno Schneider, Georg Rehm, and Bela Gipp. 2019. Enriching BERT with Knowledge Graph Embeddings for Document Classification. In *Proceedings of the 15th Conference on Natural Language Processing, KONVENS 2019, Erlangen, Germany, October 9-11, 2019*. German Society for Computational Linguistics & Language Technology, 307–314. https://corpora.linguistik.uni-erlangen.de/data/konvens/proceedings/papers/germeval/Germeval_Task1_paper_3.pdf

[50] Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. 2016. From Freebase to Wikidata: The great migration. In *Proceedings of the 25th International Conference on World Wide Web (WWW'16)*. ACM, 1419–1428.

[51] Alessandro Piscopo and Elena Simperl. 2018. Who models the world? Collaborative ontology creation and user roles in Wikidata. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 141:1–141:18.

[52] Eric Prud'hommeaux, Jose Emilio Labra Gayo, and Harold Solbrig. 2014. Shape expressions: an RDF validation and transformation language. In *Proceedings of the 10th International Conference on Semantic Systems*. ACM, 32–40.

[53] Jonathan Raiman and Olivier Raiman. 2018. DeepType: Multilingual Entity Linking by Neural Type System Evolution. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI'18)*. AAAI Press, 8 pages.

[54] Eric S. Raymond. 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly Media, Sebastopol, CA, USA.

[55] Florian Reitz. 2018. *External identifiers in dblp*. Schloss Dagstuhl – Leibniz Center for Informatics. Retrieved 14 Nov 2022 from https://blog.dblp.org/2018/10/12/external-identifiers-in-dblp

[56] Pavel Richter. 2020. *Die Wikipedia Story - Biografie eines Weltwunders*. Campus, Frankfurt, Germany.

[57] Cezar Sas, Meriem Beloucif, and Anders Søgaard. 2020. WikiBank: Using Wikidata to Improve Multilingual Frame-Semantic Parsing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, 4183–4189. https://aclanthology.org/2020.lrec-1.515

[58] Mathias Schindler and Denny Vrandečić. 2009. Introducing new Features to Wikipedia: Case Studies for Web Science. In *Proceedings of the 1st International Conference on Web Sciences (WebSci'09)*, Jim Hendler and Helen Margetts (Eds.). Web Science Trust, 56–61.

[59] Mathias Schindler and Denny Vrandečić. 2011. Introducing New Features to Wikipedia: Case Studies for Web Science. *IEEE Intelligent Systems* 26, 1 (2011), 56–61. https://doi.org/10.1109/MIS.2011.17

[60] Toni Schindler. 2020. *Accessing the World with the Wolfram Language: External Identifiers and Wikidata*. Wolfram. Retrieved 14 Nov 2022 from https://blog.wolfram.com/2020/07/09/accessing-the-world-with-the-wolfram-language-external-identifiers-and-wikidata

[61] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. 2006. The Semantic Web revisited. *IEEE Intelligent Systems* 21, 3 (2006), 96–101.

[62] Jackie Shieh. 2022. *Smithsonian Libraries and Archives & Wikidata: Using Linked Open Data to Connect Smithsonian Information*. Smithsonian Libraries and Archives. Retrieved 14 Nov 2022 from https://blog.library.si.edu/blog/2022/01/19/smithsonian-libraries-and-archives-wikidata-using-linked-open-data-to-connect-smithsonian-information

[63] Elena Simperl, Fabian Flöck, and Denny Vrandečić. 2011. Towards a diversity-minded Wikipedia. In *Proceedings of the ACM 3rd International Conference on Web Science 2011*. ACM, 1–8.

[64] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 697–706.

[65] Aparna Surendra. 2020. *An Александр by any other name*. OCCRP Team. Retrieved 14 Nov 2022 from https://medium.com/occrp-unreported/an-%D0%B0%D0%BB%D0%B5%D0%BA%D1%81%D0%B0%D0%BD%D0%B4%D1%80-by-any-other-name-819525c82d8

[66] Tabea Tietz, Jörg Waitelonis, Mehwish Alam, and Harald Sack. 2020. Knowledge Graph based Analysis and Exploration of Historical Theatre Photographs. In *Proceedings of the Conference on Digital Curation Technologies (Qurator 2020), Berlin, Germany, January 20th - 21st, 2020 (CEUR Workshop Proceedings, Vol. 2535)*, Adrian Paschke, Clemens Neudecker, Georg Rehm, Jamal Al Qundus, and Lydia Pintscher (Eds.). CEUR-WS.org, 9 pages. http://ceur-ws.org/Vol-2535/paper_7.pdf

[67] Erik M. van Mulligen, Erik Möller, Peter-Jan Roes, Marc Weeber, Gerard Meijssen, and Barend Mons. 2006. An Online Ontology: WiktionaryZ. In *KR-MED 2006, Formal Biomedical Knowledge Representation, Proceedings of the Second International Workshop on Formal Biomedical Knowledge Representation: "Biomedical Ontology in Action" (KR-MED 2006), Collocated with the 4th International Conference on Formal Ontology in Information Systems (FOIS-2006), Baltimore, Maryland, USA, November 8, 2006 (CEUR Workshop Proceedings, Vol. 222)*, Olivier Bodenreider (Ed.). CEUR-WS.org, 31–36. http://ceur-ws.org/Vol-222/krmed2006-p04.pdf

[68] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledge base. *Commun. ACM* 57, 10 (2014), 78–85.

[69] Denny Vrandečić. 2021. Building a Multilingual Wikipedia. *Commun. ACM* 64, 4 (mar 2021), 38–41. https://doi.org/10.1145/3425778

[70] Denny Vrandečić, Varun Ratnakar, Markus Krötzsch, and Yolanda Gil. 2011. Shortipedia: Aggregating and Curating Semantic Web Data. *Journal of Web Semantics* 9, 3 (2011), 334–338. Invited paper for Semantic Web Challenge 2010 finalist (3rd place Open Track).

[71] Denny Vrandečić. 2009. Towards Automatic Content Quality Checks in Semantic Wikis. In *Social Semantic Web: Where Web 2.0 Meets Web 3.0 (AAAI Spring Symposium 2009)*, Mark Greaves, Li Ding, Jie Bao, and Uldis Bojars (Eds.). Springer, 2 pages.

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

[73] Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi L Griffith, Kristina Hanspers, Henning Hermjakob, Toby S Hudson, Kevin Hybiske, Sarah M Keating, Magnus Manske, Michael Mayers, Daniel Mietchen, Elvira Mitraka, Alexander R Pico, Timothy Putman, Anders Riutta, Nuria Queralt-Rosinach, Lynn M Schriml, Thomas Shafee, Denise Slenter, Ralf Stephan, Katherine Thornton, Ginger Tsueng, Roger Tu, Sabah Ul-Hasan, Egon Willighagen, Chunlei Wu, and Andrew I Su. 2020. Science Forum: Wikidata as a knowledge graph for the life sciences. *eLife* 9 (mar 2020), e52614. https://doi.org/10.7554/eLife.52614

[74] Jay Wacker. 2017. *Announcing Wikidata References on Topics*. Quora. Retrieved 14 Nov 2022 from https://quorablog.quora.com/Announcing-Wikidata-References-on-Topics

[75] WDQS Search Team. 2022. *WDQS Backend Alternatives Working Paper*. Technical Report v1.1, 29 Mar 2022. Wikimedia Foundation, San Franciscio, CA, USA. https://www.wikidata.org/wiki/File:WDQS_Backend_Alternatives_working_paper.pdf

[76] Omer Faruk Yalcin. 2021. *Measuring and Modeling the Dynamics of Elite Political Networks*. Ph. D. Dissertation. Pennsylvania State University, University Park, PA, USA.