

Distr.
GENERAL

18 June 2012

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

**Work Session on the Communication of Statistics
(Geneva, Switzerland, 27-29 June 2012)**

Session 3: New Ways of Engaging Audiences

Wikidata – Social Media meets Open Data

Prepared by by Christoph Bergmann and Daniela Schuengel, Germany

During the last decade the way we consume and spread information changed fundamentally. With the phenomenon called “Web 2.0”, we associate a variety of web-technologies that established new individual publication channels and new ways of communication and interaction. Being one of the biggest players in the Web 2.0, Wikipedia revolutionised the availability of knowledge as a free online encyclopaedia. The main characteristic of Wikipedia is its collaborative approach: Everyone can contribute to the collection of the world’s knowledge. With Wikidata, a new platform is about to apply the collaborative approach to the collection of data. We don’t know yet, if Wikidata marks the beginning of a democratisation of data dissemination. Nevertheless, we’d like to explain the concept and how it can be useful for our purposes.

Who are the people behind Wikidata?

Wikidata is the first new Wikimedia project since the inception of Wikiversity in 2006. Same as Wikipedia it is carried by the Wikimedia Foundation¹, in this case under the responsibility of the German chapter Wikimedia Deutschland². The Project is funded by the Allen Institute for Artificial Intelligence [AI]², the Gordon and Betty Moore Foundation and Google, Inc. After the initial development the Wikimedia Foundation will take responsibility for further improvement and maintenance. So Wikidata is not intended as a German but a multilingual platform.

What is Wikidata?

Wikidata is “a free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, the other Wikimedia projects, and well beyond that.”³

¹ http://wikimediafoundation.org/wiki/Our_projects (retrieved: 04/30/2012).

² Founded in 2004, Wikimedia Deutschland today is the biggest of 40 Wikimedia chapters worldwide.

³ <http://meta.wikimedia.org/wiki/Wikidata/Introduction> (retrieved: 04/30/2012).

That means that all data published on Wikidata is under a free licence and can be reused in a variety of different ways. It is entered and maintained by Wikidata editors, who “decide on the rules of content creation in Wikidata”.⁴ As stated before, it will be completely multilingual. All components will be localised so editing and using or reusing the data will be possible in any supported language. New data will be available in every language at the same time.

What is new about Wikidata? Other Wikimedia projects provide media files (Wikimedia Commons) or articles (Wikipedia). Wikidata will provide structured data. By consequence it is stored in machine-readable form. That simplifies the reuse of all items in the database and opens up an infinite variety of possibilities of how the data can be used.

As a secondary database, Wikidata will provide – besides the statements – the origin, thus source of the data. That means that it will be possible to enter many different statements, each with the associated source⁵.

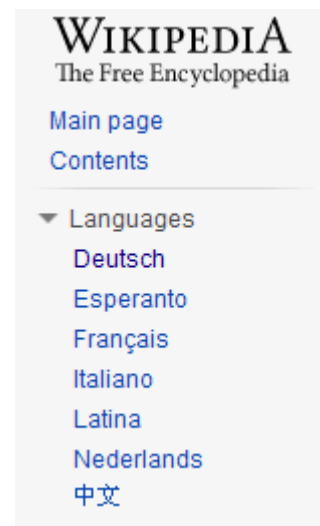
What does Wikidata?

As mentioned above, one of the primary goals of Wikidata is to support other Wikimedia projects, especially Wikipedia. The initial development will proceed in three steps that will all enhance the way articles on Wikipedia can be edited and managed.

A. Phase 1: Interwiki links

At first, Wikidata will provide an alternative to the current interlanguage link system⁶. What does this mean? At the moment there exist Wikipedias in 285 different languages⁷. Most articles are linked to other Wikipedias, treating the same subject in other languages (see picture on the right). These connections are more or less managed individually by hand for every language.

Wikidata wants to set up one page for every entity described by an article in one of the Wikipedia language editions. This page contains links to all articles describing the given entity in one of the different language editions. In the end there will be one Wikidata page for every entity and several Wikipedia articles describing the entity: one article for each language. Besides the links to the articles, the Wikidata page will contain labels and short descriptions for each supported language. Finally it will be possible to add aliases for each entity.



⁴ Ibid.

⁵ <http://blog.wikimedia.de/2012/04/27/addressing-feedback-about-wikidata/> (retrieved: 04/30/2012).

⁶ http://meta.wikimedia.org/wiki/Wikidata/Technical_proposal (retrieved: 04/30/2012).

⁷ http://meta.wikimedia.org/wiki/List_of_Wikipedias (retrieved: 04/30/2012).

Bern		
Country	Switzerland	
Canton	Bern	
District	Bern-Mittelland administrative district	
	 46°57'N 7°27'E	Details
Population	133,920 (Dec 2010) ^[1]	
- Density	2,595 /km ² (6,722 /sq mi)	
Area	51.82 km ² (19.93 sq mi) ^[2]	
Elevation	542 m (1,778 ft)	
- Highest	864 m - Gurten	
- Lowest	480 m - Aare	

B. Phase 2: Infoboxes

After setting up the Wikidata pages, phase 2 will enable the Wikidata community to add facts to the entities. These facts can be a) typed as links between entities or b) property – value pairs. Facts can be enriched with sources and other qualifiers. “The data will be fully exported in different formats, especially RDF and JSON. Wikidata will provide an API to edit the extended content.”⁸ The Wikipedia software (MediaWiki) will contain an extension allowing to include Wikidata items to infoboxes by entering a piece of code. As Wikidata

is fully multilingual, it will be possible for smaller Wikipedias to use the information of more active Wikipedia communities. For all participating Wikipedias it will take less efforts to keep the articles up to date.

C. Phase 3: Lists

In Wikipedia editors have the possibility to enrich articles using lists. Phase 3 will make it much easier to create lists that need even less maintenance in order to provide information that is always up to date.

Using more complex queries will make it possible to create lists from Wikidata. For example an editor lets Wikidata create an automatic list of all capital cities with female majors, sorted by population. Every article in any language edition using that list will automatically be updated if the source in Wikidata gets changed.

List by the CIA World Factbook (2000–2011) ^[4]

Rank 	Country/Region 	GDP (millions of US\$) 	Year 
	<i>World</i>	70,160,000	2011
	 <i>European Union</i>	17,720,000	2011 est.
1	 <i>United States</i>	15,060,000	2011 est.
2	 <i>China</i>	6,989,000	2011 est.
3	 <i>Japan</i>	5,855,000	2011 est.
4	 <i>Germany</i>	3,629,000	2011 est.
5	 <i>France</i>	2,808,000	2011 est.
6	 <i>Brazil</i>	2,518,000	2011 est.
7	 <i>United Kingdom</i>	2,481,000	2011 est.
8	 <i>Italy</i>	2,246,000	2011 est.
9	 <i>India</i>	1,843,000	2011 est.
10	 <i>Russia</i>	1,791,000	2011 est.

How can we use it?

A. Repository

Wikidata can serve as a central repository for free data, just as Wikipedia is the biggest collection of free knowledge in the web. It opens up new ways of disseminating data: The information can be added to one central database and at once it is possible to refer to it in any Wikipedia article in any supported language. Why is this so useful? In most cases Wikipedia articles are among the first results in search engines like Google. With little effort our data can become part of a huge number of such prominent articles.

As it is planned to provide sources with the data, it will be possible not only to promote data, but the original database as well. That will increase traffic on our own websites and improve our own websites' ranking on Google.

⁸ http://meta.wikimedia.org/wiki/Wikidata/Technical_proposal (retrieved: 04/30/2012).

But not only will Wikimedia projects benefit from Wikidata: As everything will be stored in machine-readable form, all third-party developers have access to the database. So without any additional costs our data can become part of other websites and apps developed with the Wikidata-API.

Just one example: Google is about to enhance its search engine with semantic features and will sooner or later answer questions directly on its own site. Other applications like Apple's Siri work with similar techniques. Therefore databases with machine-readable content are needed. Wikidata has the potential to become one of the most important sources for applications in the semantic web.

From a technical point of view it will be important to stay up to date with the development of the project. There will be an API to enter data automatically. That means Wikidata is not only machine-readable but also machine-editable.

It should be mentioned that Wikidata will not integrate data from remote sites automatically. But it is up to third-party projects to develop bots or other services that help aligning the content of Wikidata with other databases.

B. Source

It has been said that the data can be used in different ways within the different Wikimedia projects and it will be able to serve as a source for third-party projects. It will be possible to combine our own data with additional information from Wikidata. This opens up a huge variety of possibilities for new applications on the web.

Why is it “social”?

Having said that there will be an API for entering data, it is important to notice that it is still – same as on Wikipedia – the community's responsibility to decide whether data is useful or not. Consequently the Wikidata developers highlight the fact that there has to be a working community before bigger donations of data can be made.⁹

As stated before it will be a collaborative project. It will be possible to enter “competing” data from different sources. The power of one central repository stokes fears for vandalism. There will be mechanisms to restrict the access for editors of Wikidata.¹⁰ In order to sustain one's position it will help to be part of the community from the beginning. It won't be enough to develop APIs and upload data.

When will it launch?

The German chapter of Wikimedia started working in April 2012. With a very tight schedule they hope to finish development within one year. Phase 1 – the interwiki links – should be finished by this year's northern summer. Entering data and reusing it will be possible from autumn / early winter 2012. Creating lists (phase 3) is scheduled to be finished by the end of March 2013.

⁹ <http://meta.wikimedia.org/wiki/Wikidata/Introduction> (retrieved: 04/30/2012).

¹⁰ <http://blog.wikimedia.de/2012/04/27/addressing-feedback-about-wikidata/> (retrieved: 04/30/2012).

Where do we get further information?

Meta-Wiki: <http://meta.wikimedia.org/wiki/Wikidata>
Wikidata blog posts: <http://blog.wikimedia.de/tag/wikidata/>

All screenshots are taken from wikipedia.org.