

Data about living people on Wikipedia, Wikidata and Wikimedia Commons



1. Background

The Wikimedia movement is a worldwide volunteer movement aiming to make knowledge freely accessible to everyone.

This is done through a number of projects ("websites"), of which the internet encyclopedia Wikipedia is best known. Wikidata is a relatively new Wikimedia project: a freely accessible and editable database, rapidly growing in size and relevance. Wikimedia Commons is an image and sound library.

Characteristic for all the Wikimedia projects is that all information they contain is freely reusable, and all content can be edited or supplemented by anyone.

The domain names of the projects, and the servers on which they are hosted, are owned by the Wikimedia Foundation, an American non-profit. Although content management of the projects is entirely in the hands of volunteers, the Wikimedia Foundation has ultimate legal responsibility. For example, if any of the projects contain copyrighted material, the Wikimedia Foundation will ultimately be summoned by a (US) judge to remove it. However, in most cases the voluntary editors will have already removed any such material long before it comes to that.

In several countries there are Wikimedia associations (chapters) that support the Wikimedia volunteers and the Wikimedia projects in their region. These chapters have no formal powers or responsibilities regarding the content of the Wikimedia projects.

The Board of the Wikimedia Foundation has set the framework for including information or media related to living people on the Wikimedia projects:

“The Wikimedia Foundation Board of Trustees urges the global Wikimedia community to uphold and strengthen our commitment to high-quality, accurate information, by:

- Ensuring that all projects in all languages that describe or show living people have policies in place calling for special attention to the principles of neutrality and verifiability in those articles;

- Taking human dignity and respect for personal privacy into account when adding or removing information and/or media, especially in articles or images of ephemeral or marginal interest;

Investigating new technical mechanisms to assess contributions , particularly when they affect living people, and to better enable readers to report problems;

Treating any person who has a complaint about how they are portrayed in our projects with patience, kindness, and respect, and encouraging others to do the same.”

https://foundation.wikimedia.org/wiki/Resolution:Media_about_living_people

Within the Dutch Wikimedia community, questions have been raised about the ethical and legal boundaries of collecting and publishing data about living persons on various Wikimedia projects. This mainly concerns Wikipedia and Wikidata, and to a lesser extent Wikimedia Commons.

2. Data on living persons on Wikipedia

[Example: Wikipedia article on living person](#)

The internet encyclopedia Wikipedia makes knowledge accessible through encyclopedic entries written by volunteers. Wikipedia is not thematically limited and contains biographical articles about living persons.

Every Wikipedia article must meet a number of criteria: for example, the subject must be ‘notable’ (encyclopaedically relevant) and the information must be based on verifiable and reliable sources. An article must also be neutral in tone and content.

In addition, specific stricter guidelines apply to articles about living persons.

On the Dutch language Wikipedia this includes:

“Users should exercise special care when recording and editing information about living persons. Assessing and neutrally presenting such information requires a great deal of tact and empathy. The applicable laws must be strictly adhered to, and even more strictly than in other articles, these three pillars of the Wikipedia philosophy apply: neutral position, verifiability and no original research. ”

“An important rule of thumb when displaying and editing information about existing individuals is” do no harm, “that is /.. / All information about living persons must be written with restraint and detachment, and with the highest regard for the privacy of the subject. ”

Wikipedia biographical articles may contain information on matters generally perceived as personal or sensitive, such as religion, sexual orientation, illness, criminal record. Which personal data is included in an entry about a living person depends on:

- the data publicly available from a source deemed reliable by the Wikimedia community.
- applicable rules and agreements on Wikipedia
- considerations of individual Wikimedians.

This information can also be included in the so-called *infobox*, in which important facts from the biography are listed as keywords.

In addition to biographical articles about living persons, Wikipedia also contains the so-called "list articles", some of which also list living persons with a certain characteristic:

- [List of participants in the Paralympic Games](#)
- [List of atheists](#)

Finally, all articles (including biographical articles) are grouped in one or more categories that characterize an important aspect of the person's encyclopedic relevance, such as "[Dutch investor](#)", "[Transgender person](#)", "[Ex-Muslim](#)" or "[Dutch vlogger](#)".

3. Data about living persons on Wikidata

[Example: Wikidata entry](#) about a living person

Wikidata is a freely accessible, man- and machine-readable database that collects, structures and makes accessible data available in the public domain. Wikidata is not thematically limited and contains information about living persons.

The criteria for including a data item about a person in Wikidata are less strict than for including a biographical entry in Wikipedia. Not every alderman from a small municipality will qualify for a Wikipedia article, but there is no objection to including all aldermen in Wikidata.

<https://www.wikidata.org/wiki/Wikidata:Notability/en>

The data is collected from various sources and can be continuously supplemented and updated.

Wikidata also has guidelines for handling data about living persons:

“Living individuals with records in Wikidata are for the most part not famous or celebrities; their privacy should be respected. Values for living individuals should generally not be supplied unless they can be considered widespread public knowledge or are openly supplied by the individual themselves (otherwise hidden supporting references are not sufficient). As an example, the fact that someone's address is accessible by looking at a domain name registration doesn't imply that it's considered widespread public knowledge for the sake of this policy.”

https://www.wikidata.org/wiki/Wikidata:Living_people

[Properties](#) are assigned to a wiki data item about a person: descriptive characteristics. These can cover all aspects of human existence: place of birth, education, language skills, awards, employers, criminal record, family relationships, religion, political beliefs, sexual orientation, publications, offices and functions ... It is not mandatory for every biographical data item to add all possible properties, nor is it explicitly intended to add as many as possible. However, when a data item has been created because someone has become an alderman, it is possible for anyone to add information about their religion and sexual orientation in addition to information directly related to the political career - if that information is in the public domain, is widely known, and supported by sources believed to be reliable.

Which personal data is included depends on:

- the data that is publicly available and backed by a reliable source. Data sets from the government or a respectable knowledge institution, and information in serious publications are considered a reliable source.
- relevant rules and regulations on Wikidata
- considerations of individual Wikimedians.

A number of properties are seen by the Wikidata community as potentially privacy sensitive (selection)

Property	
religiion (P140)	date of birth (P569)
etnnic group (P172)	address (P6375)
sexual orientation (P91)	e-mail address (P968)
medical condition (P1050)	phone number (P1329)
convicted of (P1399)	Wikimedia-username (P4174)
lifestyle (P1576)	websiteaccount op (P553)
blodtype (P1853)	permanent resident of (P5389)
weight (P2067)	net worth (P2218)
social class (P3716)	Keybase username (P6837)
partner (P451)	checksum (P4092)
preferred pronoun (P6553)	vehicle identification number (P6322)

https://www.wikidata.org/wiki/Wikidata:WikiProject_Properties/Wikidata_properties_that_may_violate_privacy

https://www.wikidata.org/wiki/Wikidata:WikiProject_Properties/Wikidata_properties_likely_to_be_challenged

In addition, on Wikidata it is also possible to indicate family relationships between persons for whom a data item exists.

4. Data about living persons on Wikimedia Commons

[Example: Wikimedia Commons category + personal data](#)

Wikimedia Commons is the media library of Wikimedia. Wikimedia Commons contains image and sound files, including images/recordings of living persons. Wikimedia Commons has a policy regarding the inclusion of material that identifies people:

https://commons.wikimedia.org/wiki/Commons:Photographs_of_identifiable_people.

This policy was confirmed by the Board of the Wikimedia Foundation.

https://meta.wikimedia.org/wiki/WMF_Resolutions/Images_of_identifiable_people

For some individuals, Wikimedia Commons also contains an infobox of personal information from Wikidata. In addition, it has recently become possible to describe images on Wikimedia Commons in more detail by adding properties from Wikidata, e.g. 'depicted persons', 'location', 'event', even 'clothing'. The image library is also searchable on these properties.

5. Differences between information about living persons in Wikipedia and Wikidata

- Due to the stricter criteria for relevance, Wikipedia contains information about a smaller number of living persons than Wikidata
- the information in Wikipedia and in Wikidata is not necessarily identical. Wikidata may contain information not listed in Wikipedia, and vice versa.
- the information about living persons in Wikipedia is more contextualized than that in Wikidata
- the ability for readers to search specifically for individuals who meet certain criteria (religion = X, and sexual orientation = Y, and place of residence = Z) is much greater

in Wikidata than it is in Wikipedia. Search results depend on the extent to which the relevant properties have been added to data items about living persons.

6. Legal and ethical aspects

Lawyers consulted by WMNL agree that the collection of data as it currently happens in Wikidata and Wikipedia is in principle permitted under the EU GDPR. The data collected is already available in the public domain and Wikipedia / Wikidata is assumed to be covered by the GDPR exemptions as regards data collection for scientific work, archival purposes or journalism. https://nl.wikimedia.org/wiki/Wikimediaprojecten_en_de_AVG

According to a recent discussion, there are broadly speaking two positions within the Wikimedia community on the desirability of the current method of data collection:

- collecting (sensitive) data about living persons is not problematic as long as these data were already public and good sources can be referred to
- collecting (sensitive) data about living persons *from different public sources and over many years* is problematic because it creates a new level of information about the person, while the searchability of Wikidata in particular facilitates profiling and targeting.

In addition, the question is raised whether such data collection contributes to the mission of the Wikimedia movement.

7. Questions

Question 1:

On Wikidata and on Wikipedia, and to a lesser extent also on Wikimedia Commons, data about living persons is collected, structured and made available worldwide - also for reuse. This includes also information that is usually considered "sensitive" or "very personal." All of this information has been previously published in various sources deemed reliable by the Wikimedia community.

The personal data may be:

- incorporated in the text of encyclopedic articles
- have been added to an article as a category (keyword)
- included in a data item or info box,
- used to create a list of people with the same personal attribute.

Is this permitted under national and international laws and regulations?

Question 2:

If this is allowed, is there - based on e.g. jurisprudence or ongoing political and social discussions - reason to expect that the collecting and making available personal data in this way will be subject to additional rules in the near future?

Question 3:

Is this systematic collection, structuring and making available of (personal) information about living persons on Wikipedia and Wikidata in accordance with the Terms of Use for the Wikimedia sites, and other relevant guiding policies of the Wikimedia Foundation?

Question 4:

On the basis of existing insights and experience, can it be estimated whether the systematic collection, structuring and permanent availability of personal data (which in themselves were already public) could lead to adverse consequences for the persons involved? Could you indicate which data would cause the greatest risk?

Question 5:

In general, do you see a need to change the policy on handling personal data on Wikipedia and Wikidata?

Question 6:

Would you like to give the Wikimedia community advice based on your expertise and experience?