

Abstract Wikipedia/Licensing discussion

< [Abstract Wikipedia](#)

Abstract Wikipedia is a long term project with a goal of combining content from new and existing Wikimedia projects to enable contributors to create and maintain Wikipedia articles independent of language. Abstract Wikipedia will be built using contributor-created software functions and content from other Wikimedia projects, primarily Wikidata, to generate text in multiple languages.

Wikifunctions, like other Wikimedia projects, will rely heavily on contributor-submitted content. Wikifunctions will also rely heavily on contributor-driven decisionmaking around all of the projects' policy development. Therefore we should decide together which licenses to use for the components of Abstract Wikipedia.

The Abstract Wikipedia team has asked the Wikimedia Foundation's Legal department for an opinion on which licenses would be an acceptable option for each layer within Wikifunctions, and what other legal risks should be considered in choosing a license. This document is based on their recommendation and guidance.

This page first describes what these different components are. Then it describes which licenses make sense for which component. This is followed by a discussion of the different options and their interactions with each other. The page is closed by statements of individual opinions.

Overview and request

All contributions to Wikifunctions and the wider Abstract Wikipedia projects will be published under free licenses. We recommend to publish Abstract Content under the same license as the content of Wikipedia, i.e. CC BY-SA. We further recommend to publish textual content on Wikifunctions under the same license, to publish function signatures under CC-o, and function implementations under the Apache license.

We would like to know whether you support these recommendations, or whether you have other suggestions, feedback, concerns, etc. Please [comment on the talkpage](#).

Abstract Wikipedia

(Discussion)

General

[Overview](#) — [Vision](#)

[Glossary](#)

[Function model](#)

Updates

Development plan

[Summary](#) — [Name](#)

[Goals](#)

[Organization](#) — [Requirements](#)

[Architecture](#) — [Components](#)

[Tasks](#) — [Implementation phases](#)

(single page version of all above)

Notes, drafts, discussions

[Overview of notes](#)

- [Pre-generic function model](#)
- [Object creation requirements](#)
- [First evaluation engine](#)
- [Function evaluator call](#)
- [Representation of dates](#)
- [Representation of languages](#)
- [Representation of errors](#)
- [Error messages on ZObjects](#)
- [Reserved ZIDs](#)
- [Canonical and normal](#)
- [User stories](#)

[Discussion topics](#)

[External outreach](#)

[Ideas](#)

[Licensing discussion](#)

[Related and previous work](#)

- [Natural language generation](#)
- [Theories of semantics and meaning representations](#)

– [Workshops](#)

[ZObject](#)

Contents

[Overview and request](#)

[Components](#)

Example of the components

Legal analysis and copyright standards by type of content

Facts

Software

Text or media content

Recommendation from the WMF legal department

Function Signatures

Function Implementation

Abstract Content

Output Content

Recommendations by the development team

Request for input

Examples & mockups

Examples

Examples of functions

Mockups

Example: Jupiter

Examples: Function composition

Data tools

Data


Lexicographical coverage

Historical

Papers, press, and videos

Historic proposal

Naming contest—Logo contest

 (https://meta.wikimedia.org/w/index.php?title=Special:Translate/page-Template:Abstract_Wikipedia_nav_box&language=en&action=page)

Archives

Components

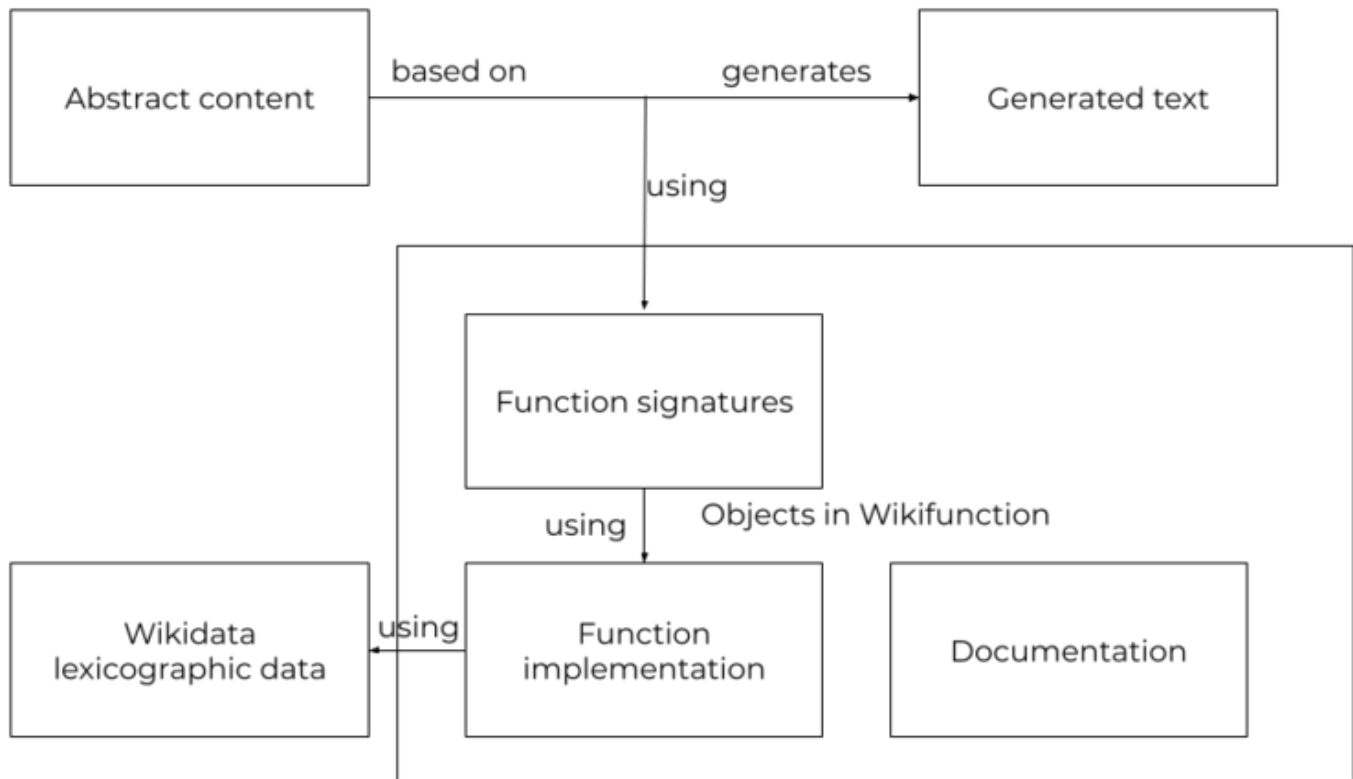
An important step towards Abstract Wikipedia is Wikifunctions, a wiki of software functions that can be combined with items from Wikidata. Wikifunctions will be a software hosting platform, allowing users to write and execute code from their web browser. Wikifunctions will allow users to submit and execute functions that can incorporate content from other Wikimedia projects.

For the sake of licensing, we discuss content at four levels:

- **Function Signatures:** The definition of a function, i.e. its name, number and types of input arguments, and the type of the output;
- **Function Implementation:** The code that is called and executed within Wikifunctions, which may include contributor-submitted code and libraries available on the platform;
- **Abstract Content:** The abstract representation of text or text fragments, which is essentially a set of specialized function calls to produce an output; and
- **Output Content:** The text produced through calling Function Implementations on Abstract Content, and often pulling in Wikidata Content.

An example of each is provided in the section below this.

The following image sketches out the different components of Abstract Wikipedia relevant for the discussion. We describe the different components in the subsequent text.



Wikifunctions will consist of function signatures, function implementations, and other objects. These other objects can be of various types, and the set of types is extensible. Objects will have documentation.

Function definitions are like APIs, they give the name of the function and arguments as well as the types of the arguments and of the result. **Implementations** are the source code of the functions - they tell the computer how to turn the arguments into an answer. **Documentation** can be made for any definition, implementation, and any other object in Wikifunctions. Other objects in Wikifunctions could be testers, individual strings, lists, types, abstract content, etc.

Some of the functions will be functions that take **abstract content** and generate **output content** (often natural language text) out of it. Some of the functions will access data in Wikidata or other locations and generate natural language text out of it. The functions may use lexicographic data from Wikidata in order to generate the text.

The generated text might be integrated into each Wikipedia to fill knowledge gaps.

Example of the components

Note: These examples are highly simplified, and show example values in English which would technically be Wikidata QIDs for internationalization ([examples](#)).

Given a constructor Superlative with the keys subject, quality, class, and location constraint, we can have the following **abstract content**:

```

Superlative(
  subject: Jupiter,
  quality: large,
  class: planet,
  location constraint: Solar System)

```

In Wikifunctions, we would have the following **function definition**:

```
generate_text(superlative, language) : text
```

I.e. a function that takes a superlative object (as given in the abstract content above) and a language (such as English) and returns a text.

This could be then a possible **function implementation** in Python:

```
def generate_text(superlative, language):
    if language == English:
        subject = get_label(superlative.subject, language)
        adjective = superlative_form(superlative.quality, language)
        class = singular_form(superlative.class, language)
        location_clause = make_location_clause(superlative.location_constraint, language)
        text = '\'.join(subject, 'is the', adjective, class, location_clause)
        return text.capitalize_first(text)
    if language == Hausa:
        ...
    ...
```

The application of the function to the abstract content would result in the following **output content**:

(in English) *Jupiter is the largest planet in the Solar System.*

(in Croatian) *Jupiter je najveći planet u Sunčevom sustavu.*

This text can then be shown by a language edition of Wikipedia in order to provide a common baseline of knowledge about Jupiter.

Legal analysis and copyright standards by type of content

The Wikimedia movement has adopted a general licensing policy that favors free culture licenses and open source software licenses. The Creative Commons Zero (<https://creativecommons.org/publicdomain/zero/1.0/>) (CC0) deed provides a waiver that aims to ensure content is in the public domain (or a jurisdiction's local equivalent to the public domain) around the world. The Creative Commons Attribution-ShareAlike (CC BY-SA) license enables people to use content by complying with minimal attribution requirements and ensuring that subsequent adaptations are released under the terms of a compatible license. Finally, for software, Wikimedia follows the list of approved licenses (<https://opensource.org/licenses>) from the Open Source Initiative (OSI).

Facts

Facts themselves are generally not protectable under copyright law. In the United States, the Supreme Court described this principle in *Feist Publications Inc. v. Rural Telephone Service Co.*: “The most fundamental axiom of copyright law is that '[n]o author may copyright his ideas or the facts he narrates.’” This is a basic element of copyright law that applies regardless of whether content is explicitly released under a license.

The Wikimedia projects have adopted CC0 for projects that are designed to collect basic facts. For example, this includes items or properties on Wikidata, or structured data on Wikimedia Commons. CC0 provides a basic, international waiver of copyright to ensure that factual content is free to be used without restriction. This makes it easier for content to be discovered, reused, and cited elsewhere. This license may also incorporate unprotectable facts that are copied from sources under other licenses, such as text from Wikipedia that is otherwise available under CC BY-SA.

Software

Software is generally protectable under copyright law, although not all aspects of software should be treated the same. Wikimedia has taken the position that the organization and basic function of APIs are not copyrightable as a matter of US law.

For copyrightable software, the Foundation adopts an open source software license to allow it to be freely reused under minimal conditions. Wikimedia's guiding principle on freedom and open source provides that the Foundation should release all of the code it creates under an applicable open source license. In areas where Wikimedia supports user-created software, such as Wikimedia Cloud services, it also requires that the software be released under an open source license. There are a wide variety of licenses that qualify as open source, so to clarify this requirement Wikimedia typically turns to the list of OSI-approved licenses for software.

For MediaWiki, the primary software license is the GNU General Public License (version 2.0 or later). For other projects, Wikimedia usually adopts a more permissive license, such as the Apache License (version 2.0) which has been used for individual projects, extensions, and mobile applications.

Text or media content

Under the Wikimedia content licensing policy, Wikimedia projects may host content that is available under a Free Content License, in the public domain (such as expired copyright or otherwise uncopyrightable content), or fair use justification (in certain limited circumstances). For most Wikimedia projects, including Wikipedia this means that text and media is available under CC BY-SA (version 3.0).

Recommendation from the WMF legal department

Wikimedia should establish the license for each type of content through clear and simple policies for Wikifunctions. Wikimedia should create documentation that explains the licenses for each type of content, including the reason for selecting this license. For software in particular, there should be guidance about only importing third-party code under an acceptable license. The user interface should include a license grant that is appropriate for each form of content.

Function Signatures

Recommendation: Function signatures should be CCo.

Function signatures should be composed of basic components (e.g., a list of parameters) that are based on underlying functionality. The actual content of the function signatures is unlikely to be protectable by copyright, or where it is protectable, it may be used under fair use in the US. CCo is an appropriate license for information that may not be eligible for copyright at all. Use of CCo avoids creating confusion or misleading reusers to believe there are copyright limitations where none exist. Additionally, CCo will ensure maximum interoperability with the other open source or free culture licenses.

Function Implementation

Recommendation: Function implementation should be under the Apache License.

Function implementation should be under an OSI-approved license. If the Wikifunctions team and the community wishes to limit this to one license, then the Apache License would provide an ideal level of permissive flexibility.

Additionally, Wikifunctions may allow other permissive licenses that are compatible with the Apache License, such as the MIT License or (3-clause) BSD License. Allowing an additional set of license options may allow users to import more content from other third-party sources. However, it would also require creating additional

software requirements, such as a user interface to select a license and display the appropriate license notices. Wikifunctions may choose a single license, for the sake of simplicity, during the initial launch and then consider adding support for multiple licenses later based on need.

Abstract Content

Recommendation: Abstract content should be licensed under CC BY-SA or CCo.

Abstract content may be released under a CC BY-SA license or any other equally permissive license that suits the project's objectives and meets Wikimedia's licensing policy. Wikimedia has significant latitude in choosing the best license for Abstract Content.

Choosing CC BY-SA would be a standard choice, and would offer the benefit of consistency with Wikipedia and most other Wikipedia projects. It would enable users to copy and incorporate Wikipedia content into Abstract Content in some way. However, it would also require preserving an edit history or some equivalent contribution history mechanism for Abstract Content, to enable people to provide attribution to the content's list of authors.

Alternatively, Abstract Content could be released under the more permissive CCo terms. This would allow the software to bypass the attribution requirements, but limit users' ability to copy or incorporate any protectable portions of Wikipedia articles or other sources.

Output Content

Recommendation: Output Content should be licensed under CC BY-SA or CCo.

Since Output Content is generated via software that combines multiple data sources, there may be questions about whether the resulting product is copyrightable at all. In 2019, the US Copyright Office requested comments about how content created by AI algorithms or processes should be handled under copyright law. In the Wikimedia Foundation's submission in response (https://www.uspto.gov/sites/default/files/documents/Wikimedia-Foundation_RFC-84-FR-58141.pdf), we explained that AI algorithms should be treated like any other software tool and that the tool's user should be considered the copyright holder. Following the same principle, Wikimedia may consider Output Content as a work of creativity by the authors of the Abstract Content. It would therefore likely be most effective if the output content is licensed consistently with existing Wikimedia projects, and can be discussed by the communities as to which of the licenses currently in use would be preferred.

Recommendations by the development team

The development team recommends to follow the recommendations by Legal, i.e. to choose CCo as the license for Function Signatures; Apache for Function Implementations (and to start with a single license, and only when we recognize the need for multiple licenses to extend Wikifunctions to support multiple licenses); to choose either CCo or CC BY-SA for the Abstract Content and the Output Content.

For documentation and other textual content of Wikifunctions we will choose CC BY-SA in order to preserve compatibility with most other Wikimedia projects regarding textual content. For other objects in Wikifunctions besides implementations, we will keep it consistent with the choice for Function Signatures.

The development team further recommends to choose CC BY-SA for Abstract Content and Output Content. Whereas one could argue that Abstract Content is more similar to the structured data of Wikidata than to the natural language text of Wikipedia, we think that there are a number of factors that make Abstract Content sufficiently similar to text:

1. Editors have a very fine-grained selection of which facts are being displayed and which are not. In Wikidata we strive for completeness over careful selection.
2. Editors have a very fine-grained control of the order the facts are being displayed in, constituting narrative elements, which are not available in Wikidata.

3. We expect the natural language generation to allow editors to express to some degree of emphasis and selection of wording.

All of these point towards Abstract Content being more similar to text than to a collection of facts, and therefore we suggest that we follow the same license that we use for the text in Wikipedia, which is CC BY-SA. On the other hand, one could argue that by putting Abstract Content under CCo we open the space for a larger amount of possible reuse in applications we cannot even imagine yet, nevermind properly decide the legal framework for. CCo most certainly allows the most freedom in the reuse of the Abstract Content.

Request for input

We would like to invite the community to **discuss these recommendations** and hopefully to find consensus around the licensing decision. The goal is to keep the discussion open for about four weeks, and, if necessary, to extend and restructure it. In case this turns out to be insufficient to reach a consensus, we might restructure the licensing discussion to focus solely on Wikifunctions for now, and then follow up with a discussion about Abstract Wikipedia.

To guide the license choice, it may be useful to consider and discuss the following questions:

1. What are the long-term objectives of the projects, and how can a copyright license support these objectives?
2. Should the people involved in creating Abstract Content receive credit?
3. How valuable is it to preserve consistency and compatibility with the licenses on Wikipedia?

The specific question to the community is whether to follow the given recommendations, or whether a different proposal would be more to the liking of the community. Particular questions of interest surround the licensing of Function Implementations, and the licensing of Abstract Content (and thus Output Content).

Even if you are just supporting the recommendations above, it would be great to see your voice expressed explicitly, in order to get a better understanding of what the community tends towards. If no rough consensus can be reached, we will organize a formal vote.

Comments are welcome in any language.

Retrieved from "https://meta.wikimedia.org/w/index.php?title=Abstract_Wikipedia/Licensing_discussion&oldid=22368268"

This page was last edited on 23 November 2021, at 08:55.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.