

Abstract Wikipedia Grant Report

Round 1: Grants given in the 2022–2023
fiscal year

General Information & Instructions

These reports will be relatively brief and at a high-level. The content should be largely future-focused. Please include a link to financial reporting in the Further Reading section. If available, please include additional documents (media coverage, impact stories, images, videos, etc) as attachments or links in the Further Reading section. The main section of the report should be less than 2 pages. There is no limit to the number of pages/amount of content in the Further Reading section.

The audience for these reports is the Endowment board and Endowment donors (past, current, and future). Donors contribute the initial funds, the proceeds of which are eventually disbursed as grants. The Endowment Board sets the grant strategy and decides how much to grant and to what projects. Please avoid detailed technical language. If there is existing, highly technical information, please include it in the Further Reading section.

The period for this report will be the previous 10 months, including the period already reported on in the November 2023 update. Please reach out to Caitlin Virtue (cvirtue@wikimedia.org) if you have any questions, need support, or would like to suggest changes to the reporting process or content.

Grant Name: Abstract Wikipedia

Grant Amount

\$1,000,000.00

Grant Summary

Abstract Wikipedia (meta.wikimedia.org/wiki/Abstract_Wikipedia) is a long-term project to enable the creation of language-independent Wikipedia articles. Abstract Wikipedia relies on **Wikifunctions**, a new Wikimedia project that allows anyone to create and maintain code. Wikifunctions provides code that translates relationships between concepts, and data from Wikidata, into natural human language, like a regular Wikipedia article.

Once more pieces of this project are in place, this will mean that any wiki – especially small to medium wikis – will be able to dramatically increase the number of articles available in their language.

Project goal for 2022–2023

Communities can build, use, and maintain a library of functions on Wikifunctions.

Why it's relevant to the Endowment

The ultimate goal of Abstract Wikipedia is encyclopedic knowledge that is language-agnostic; it represents the next evolution of Wikipedia, a technological leap to stay relevant in a digital environment centered around data and structured content. As an entirely new conceptual way of representing knowledge, Abstract Wikipedia also involves a higher degree of risk than traditional technical and product initiatives at the Foundation, all of which makes it an excellent fit for the Endowment.

November 2023 Update

After three years of development, we are excited to share that [Wikifunctions](#) is slowly beginning to roll out. Just like any other new Wikimedia initiative, we are rolling out Wikifunctions with minimal content and features to start. Logged-in contributors can run a few early functions, and editing will be limited as we test the project's stability. You can request [edit access](#) on Wikifunctions, and we will give out the rights to more people as we test and learn. Things will bend, break, get fixed, and break again as we build up the project's capabilities. For now, we are excited to announce that the Wikimedia communities have decided on the very first function: [join together](#). The function takes two strings and returns them combined. If you are logged-in, you can try it out now on [Wikifunctions](#).

WIKIMEDIA

ENDOWMENT

We are looking forward to a thriving community to grow and maintain a library of many useful and interesting functions in the future.

Further Reading

- Diff: [Wikifunctions is starting up](#)
- [Weekly Updates](#)

10 Month Report

Please answer the following questions in 2 pages or less:

Did you accomplish your goals for the year?

Yes. Our goal for the year was to launch Wikifunctions as a new top-level Wikimedia project, as a space for a new community to come together and start working and maintaining a library of functions. We set ourselves a goal to have the community create 1000 functions, 1500 implementations of the functions, and 2000 tests for the functions by the end of our financial year in June 2024.

We achieved the goal of launching Wikifunctions in general release at the end of 2023. By the end of May we crossed 1,200 functions, 1,900 implementations, and 3,400 tests, comfortably reaching our targets. The community has made more than 29,000 edits to Wikifunctions.

So far, Wikifunctions has been visited 5.8 million times, with April alone accounting for 1.76M visits. Wikifunctions was edited from more than 40 countries in April, and visited from more than 200 countries. Our community channels are active and have hundreds of members.

Why is this work important to the Wikimedia projects over the long-term?

Wikifunctions will provide a function repository that can be used from Wikipedia and the other Wikimedia projects. This aims to close knowledge gaps, particularly in languages where the communities of contributors are small.

Currently, Wikipedia is available in more than 300 language editions. In many of these, knowledge is missing or outdated. It is difficult for speakers of small language editions to contribute knowledge that can be disseminated to larger language editions. On the other hand, knowledge from a small number of language editions, most notably English, is being actively pushed into smaller wikis. Abstract Wikipedia aims to help fill these knowledge gaps, and Wikifunctions is the underlying mechanism to make this possible.

In many Wikipedia communities, the ability to call code in order to maintain and build articles is crucial. Wikipedia's infoboxes are maybe the best known example for this, but so are converting values between different units, dates between different calendars, or creating charts and tables of population figures. These are currently functionalities made available through

templates and modules. But templates and modules need to be set up in each of the language communities, and in many communities there is a severe lack of technical skill, which can lead to knowledge gaps or, even worse, a concentration of editorial power.

Wikifunctions aims to provide a common repository of functionalities and capacities that all contributors across languages can benefit from, contribute to, and maintain. This will eventually include the capability to create natural language text, and thus make large amounts of high quality encyclopedic content available for languages that previously did not have access to it.

What is, or will be, the impact of the work you did this year?

Although we are still at the very beginning, and support only a small number of data types, we have already been focusing on generating functions for a number of languages. Among them underrepresented languages such as Igbo. When talking to an Igbo community member, we were not prepared to see the excitement in their reaction: not only did they see, for the first time, a function working on Igbo, but could easily create more functions themselves, refine them collaboratively, and immediately share them with the whole world! The excitement was extremely motivating. We speak about this more in this newsletter:

https://www.wikifunctions.org/wiki/Wikifunctions:Status_updates/2024-02-01

Moving forward

What is not done yet?

There are two crucial milestones we are making progress on in the coming months:

1. Connecting Wikifunctions to Wikidata, so that we can use data from Wikidata in order to run functions. This will allow us to use the billions of facts that Wikidata contains about the world, and the tens of millions of words and word forms in Wikidata to generate natural language text.
2. Connecting Wikipedia to Wikifunctions, in order to allow for Wikifunctions to be called from Wikipedia. This would enable us to generate content through functions in Wikipedia, which are automatically updated when the content, the functions, or the data in Wikidata are updated.

Once these two tasks are done, we will unlock the potential to generate up-to-date, high-quality content in many languages.

What are the open questions? / What don't we know yet?

WIKIMEDIA

ENDOWMENT

How will these functions be welcomed by the Wikipedia communities? How many contributions will be coming from underrepresented groups, and how is that different from today? How much of the content can we represent with this approach?

What else needs to be improved/fixed in this area?

MediaWiki is in the process of modernizing in order to enable fragments of content to be individually cached and then assembled. This work is needed for achieving our next milestones, and we are in collaboration with the MediaWiki team to make this happen. We also need to support the community in developing more of a feeling of collaboration across languages. Some major communities don't often look beyond the borders of their own wikis. It would be good to create a global community that works together.

Are there external trends that are having or will have an effect on this area?

The advances in machine learning can also help us in some of the usability challenges we are expecting to face in the further development of Abstract Wikipedia, leading to an easier editing experience.

Further Reading

We recommend to look at our weekly newsletter for more details:

- https://www.wikifunctions.org/wiki/Wikifunctions:Status_updates

The following DIFF posts have also been describing our progress:

- <https://diff.wikimedia.org/2024/03/27/wikifunctions-where-we-are-now-and-what-you-can-do/>
- <https://diff.wikimedia.org/2023/08/07/wikifunctions-is-starting-up/>

Financial reporting:

Abstract Wikipedia Endowment Grant Financial Report Jul 2022 - Jun 2023

Kiwix Grant Report

Round 1: Grants given in the 2022–2023 fiscal year

General Information & Instructions

These reports will be relatively brief and at a high-level. The content should be largely future-focused. Please include a link to financial reporting in the Further Reading section. If available, please include additional documents (media coverage, impact stories, images, videos, etc) as attachments or links in the Further Reading section. The main section of the report should be less than 2 pages. There is no limit to the number of pages/amount of content in the Further Reading section.

The audience for these reports is the Endowment board and Endowment donors (past, current, and future). Donors contribute the initial funds, the proceeds of which are eventually disbursed as grants. The Endowment Board sets the grant strategy and decides how much to grant and to what projects. Please avoid detailed technical language. If there is existing, highly technical information, please include it in the Further Reading section.

The period for this report will be the previous 10 months, including the period already reported on in the November 2023 update. Please reach out to Caitlin Virtue (cvirtue@wikimedia.org) if you have any questions, need support, or would like to suggest changes to the reporting process or content.

Grant Name: Kiwix

Grant Amount

\$250,000.00

Grant Summary

Kiwix (kiwix.org) is an offline reader for online content like Wikipedia, Project Gutenberg, or TED Talks. It makes knowledge available to people with no or limited internet access. The software as well as the content is free to use for anyone.

Kiwix makes highly compressed copies of entire websites (text, images, and videos) that each fit into a single (.zim) file. Zim files are small enough that they can be stored on users' mobile phones, computers, microSD cards, or small, inexpensive mobile hotspots. Kiwix then acts like a regular internet browser, except that it reads these local copies. People with no or limited internet access can enjoy the same browsing experience as anyone else.

Project goal for 2022–2023: Improve the quality of the content and stability of the platform, notably through a new content management system that ensures the integrity of offline files.

Why it's relevant to the Endowment:

Kiwix is an example of technical innovation that sprung from the Wikimedia community. It addresses a need complementary to the Foundation's main efforts, by spreading free knowledge beyond people with fast internet access. In 2023, Kiwix had more than 10 million users in 200+ countries and territories.

Annual reports

Kiwix' fiscal year runs along the calendar year, so this grant runs across two distinct periods on our end. Both reports include a P&L report.

 [2022 Annual Report.pdf](#)

 [2023 Annual Report.pdf](#)

WIKIMEDIA

ENDOWMENT

10 Month Report

Did we accomplish our goals for the year?

Mostly yes, with one big exception: we have not been able to set up a content management system as we expected. Other than that, our technical debt on the platform side has been mostly resorbed, and we have pursued our shift to providing more content. Our user base also kept on growing, and we are now looking at 6-8 million users worldwide at the end of 2022/first half of 2023.

Why is this work important to the Wikimedia projects over the long-term?

From our perspective, the long-term problem that Wikimedia projects face is that future content and editor growth potential is highest in regions where contributors are harder to reach for a host of reasons, one of them being the lack of reliable internet connectivity. Kiwix functions as a gateway to Wikipedia, providing people with content they need and making them acquainted with Wikipedia and its sister projects.

We at Kiwix are keenly aware of the issue as one of our main avenues for new deployments are Wikimedia User Groups, mostly in Sub Saharan Africa and Latin America. The Open Foundation West Africa, the movement's Ghanaian affiliate, has also understood this very much when they've launched their Kiwix4Schools train-the-trainers program, training a first cohort of 70+ Kiwix evangelists from across the continent.

Another thing of interest is that we are a lot freer to experiment with content that the Foundation is. A couple of examples that come to mind are the mini and nopic Wikipedias that we distribute: while the latter simply is an attempt at saving bandwidth by removing images from the zim files (saving about 70% of disk space), the mini version acts upon earlier WMF research that showed that 60% of readers were content with the information contained in the lead and infobox or articles. We have pursued this venue by generating an ever increasing number of subsets based off Wikiprojects (geography, computer sciences, history, etc.) not to mention the first of them all Wikimed (medical wiki), which is now distributed in 11 different languages and counts close to half a million (connected) users. While Wikipedia functions as a one-size-fits-all behemoth, we've also shown that there is demand for more niche, snackable content.

Last but not least, we integrated 27 small and indigenous languages to our library as they got released from the Wikipedia Incubator.

What is, or will be, the impact of the work we did this year?

The most important thing is that we have, again, expanded our audience to about 8 million users worldwide. This is a conservative value as it is solely based on feedback from our partner

WIKIMEDIA

ENDOWMENT

orgs (e.g., Fondation Orange deploying Wikimedia content to 500,000 West African kids; ditto with Libraries without Borders).

Wikimedia content does not live in a vacuum and is often used in support of educational programs and other materials. This, for instance, is the case for Fondation Orange's deployments, where school attendance is seen to increase wherever they happen. Another example also happens in prisons, where Wikipedia as a whole is a no-go, but the Computer sciences subset is actively used by training programs such as the Last Mile, Persevere or Unlocked Labs.

Moving forward

We have not explored yet the full extent to which snackable content is in demand. The next logical step for us is to make it easier still for people to consume and share encyclopedic knowledge, and that will probably mean looking into custom apps *à la* Wikimed (e.g. for large sporting events like the World Cup). We will also want to see whether there is appetite for a Wikipedia-on-demand service whereby people could simply be able to generate their own subsets so as to meet specific needs.

We really, really need to fix MWoffliner. The technical debt on our Mediawiki scraper is mounting, and any change at the WMF infrastructure level has a direct impact on how/whether we can update our content.

We need to be better at measuring impact. This is a more open question, and the answer will come from the Foundation as much as from us. Everybody agrees Wikipedia is a great resource. But how great? What is the monetary lifetime value of being able to access the Sum of all Human Knowledge? We have anecdotal evidence from our users that it is an important resource for their studies, and therefore their future income. As we increase collaboration with other funders, some of which want to work on a ROI basis, we need to have a better understanding of the concrete value we bring.

Does Kiwix want to be an anti-censorship tool, and do we have a choice anyway? Downloads went through the roof in the first half of 2022 when the Russian government threatened censorship of the encyclopedia. This year again, more than 20% of our server traffic also came from mainland China, where the Internet is fast but Wikipedia is blocked. And for years we have known of North Korean defector groups smuggling Kiwix/Wikipedia on flash drives back in the country. We see an increasing trend towards censorship, but branding Kiwix as an anti-government censorship tool, at least in our very neutral Switzerland, comes with a number of risks.

Machine Learning Grant Report

Round 1: Grants given in the 2022–2023
fiscal year

General Information & Instructions

These reports will be relatively brief and at a high-level. The content should be largely future-focused. Please include a link to financial reporting in the Further Reading section. If available, please include additional documents (media coverage, impact stories, images, videos, etc) as attachments or links in the Further Reading section. The main section of the report should be less than 2 pages. There is no limit to the number of pages/amount of content in the Further Reading section.

The audience for these reports is the Endowment board and Endowment donors (past, current, and future). Donors contribute the initial funds, the proceeds of which are eventually disbursed as grants. The Endowment Board sets the grant strategy and decides how much to grant and to what projects. Please avoid detailed technical language. If there is existing, highly technical information, please include it in the Further Reading section.

The period for this report will be the previous 10 months, including the period already reported on in the November 2023 update. Please reach out to Caitlin Virtue (cvirtue@wikimedia.org) if you have any questions, need support, or would like to suggest changes to the reporting process or content.

Grant Name: Machine Learning

Grant Amount

\$950,000.00

Grant Summary

The **Machine Learning** team (mediawiki.org/wiki/Machine_Learning) at the Wikimedia Foundation runs and maintains the infrastructure that powers machine learning models on Wikimedia sites. While the models themselves are developed by communities or the Research team, the Machine Learning team makes sure the technical platform can operate those models at the scale necessary for Wikimedia sites, and in an ethical manner.

Machine learning models implement algorithms whose behavior derives from a set of training data. On Wikimedia sites, this notably involves algorithms to measure the quality of Wikipedia articles, or the likelihood that a specific edit is vandalism. Such features enable contributors to focus on editorial decisions that require more complex human judgment.

Team goals for 2022–2023

Launch a minimal viable product for [Lift Wing](#), a scalable, machine learning model and deprecate ORES (legacy infrastructure). Lift Wing is a newer, more scalable model and more appropriate for Wikipedia's current and future needs.

Why it's relevant to the Endowment

Machine learning is one of the newer technologies that enable Wikipedia and its sister sites to stay relevant in the digital environment of today and tomorrow. Lift Wing enables volunteer editors to delegate straightforward decisions to automated systems, thus freeing up their time for more complex work. It is a great example of technical innovation that supports the long-term sustainability of Wikipedia and of its community of editors.

November 2023 Update

Diff: [Exploring paths for the future of free knowledge: New Wikipedia ChatGPT plugin, leveraging rich media social apps, and other experiments](#)

[MinT: Supporting underserved languages with open machine translation](#)

https://www.mediawiki.org/wiki/Machine_Learning

10 Month Report

Please answer the following questions in 2 pages or less:

Did you accomplish your goals for the year?

Yes. The goal of the Machine Learning team for FY 2022-2023 was to launch the Foundation's new machine learning infrastructure designed to host modern machine learning models. Since the beginning of the Foundation's work on machine learning, the machine learning models hosted on Wikimedia servers used a homegrown system called ORES. This system was originally designed for the needs of researchers to explore how different models interacted with Wikipedia. However, as Wikipedia became more popular and as machine learning models became larger and more complex, there was a need to replace ORES with a modern infrastructure specifically designed to scale to the massive traffic received by Wikipedia and the Foundation's other projects. For this reason, the Machine Learning team set a goal for FY 2022-2023 to launch this new infrastructure, which is called Lift Wing. We fully accomplished this goal in FY 2022-2023. The ORES infrastructure has been shut down and Lift Wing is now hosting not only all the models previously hosted on ORES but also new, modern models that previously would not be able to be hosted on ORES.

Why is this work important to the Wikimedia projects over the long-term?

There is a significant long-term opportunity to use modern machine learning models (often called Large Language Models [LLMs]) to assist the work of editors - including detecting vandalism and writing article descriptions for human editors to review/approve. However, serving these models to millions of users is a significant engineering challenge. Lift Wing is designed from the ground up to easily handle these models at scale. Launching Lift Wing was a prerequisite for procuring and installing GPUs that are a required hardware component for LLMs, and with Lift Wing launched, we have already started procuring GPUs for that very purpose.

What is, or will be, the impact of the work you did this year?

In FY2022-2023 we could see that in a few years there would be significant opportunities to use LLMs to help editors on Wikipedia -- but that procuring and building that infrastructure would take time. Investing in launching Lift Wing gave us the infrastructure required to deploy our own LLMs at scale. Looking at the product plans for the first quarter of FY2024-2025 we can see that our estimate was correct, many product teams are planning on using LLMs for upcoming features for Wikipedia and other Wikimedia products, including:

- Logo detection for Wikimedia Commons to reduce the burden on human editors to handle copyright takedown requests.

- Using LLMs to detect potential issues with the edits of new editors (i.e. not using neutral point of view, not using reliable sources, etc.) to reduce new editor churn.
- Writing short descriptions of articles (called “beams”) and asking human editors to correct/approve.

Serving these models to millions of users cost effectively requires the infrastructure improvements made possible by launching Lift Wing.

Moving forward

We completed our goals for FY2022-2023. But this is just the beginning for Lift Wing.

Lift Wing was designed to give the Wikimedia Foundation the ability to host the most modern machine learning models. Over the last few years, we have seen the capabilities of these models increase near-exponentially. For this reason, all the external trends we see lead us to believe the opportunities for using large language models to support the work of human volunteers on Wikipedia and related products will only grow, and thanks to Lift Wing we are positioned to take full advantage of it.

Further Reading

Financial reporting

Machine Learning Endowment Grant Financial Report Jul 2022 - Jun 2023

Wikidata Grant Report

Round 1: Grants given in the 2022–2023 fiscal year

General Information & Instructions

These reports will be relatively brief and at a high-level. The content should be largely future-focused. Please include a link to financial reporting in the Further Reading section. If available, please include additional documents (media coverage, impact stories, images, videos, etc) as attachments or links in the Further Reading section. The main section of the report should be less than 2 pages. There is no limit to the number of pages/amount of content in the Further Reading section.

The audience for these reports is the Endowment board and Endowment donors (past, current, and future). Donors contribute the initial funds, the proceeds of which are eventually disbursed as grants. The Endowment Board sets the grant strategy and decides how much to grant and to what projects. Please avoid detailed technical language. If there is existing, highly technical information, please include it in the Further Reading section.

The period for this report will be the previous 10 months, including the period already reported on in the November 2023 update. Please reach out to Caitlin Virtue (cvirtue@wikimedia.org) if you have any questions, need support, or would like to suggest changes to the reporting process or content.

Grant Name: Wikidata

Grant Amount

\$1,000,000.00

Grant Summary

Wikidata (wikidata.org) was launched in 2012 to create a free knowledge base about the world that can be read and edited by humans and machines alike. It seeks to provide data in all the languages of the Wikimedia projects, and allow for the central access to data in a similar vein as Wikimedia Commons does for media files.

Currently containing over a hundred million items, Wikidata has experienced incredible growth over the past ten years. It is the most edited Wikimedia project, and has developed a dedicated community beyond Wikipedians.

WIKIMEDIA

ENDOWMENT

Project goals for 2022–2023: Empower Wikidata contributors and data users with a data quality tool, and increase use of Entity Schemas that enable editors to create and maintain data model definitions for a consistent ontology.

Why it's relevant to the Endowment:

Like Abstract Wikipedia (which will pull data from Wikidata), it represents a major step towards structured knowledge and Wikimedia's direction of "Knowledge as a Service." Wikidata can connect knowledge more easily with automated programs, voice assistants, and other interfaces of the future that don't exist yet, thus enabling Wikimedia to stay relevant in the long term. By crossing linguistic boundaries and providing integration with other Wikimedia sites, it's also a critical part of the future community infrastructure.

November 2023 Update

Point of Contact: Guillaume Paumier

10 Month Report

Please answer the following questions in 2 pages or less:

Did you accomplish your goals for the year?

Yes, the stated goals for this body of work have largely been met:

- Wikidata Platform: Empower contributors and data users - Product platform
 - Goal achieved - Data Use
 - Goal achieved - Underrepresented languages
- Wikibase Ecosystem: Empower knowledge curators to share their data
 - Goal achieved - Flagship partners
 - Goal achieved - Wikibase cloud / Flagship partners
 - Goal achieved - Wikibase.cloud

Why is this work important to the Wikimedia projects over the long-term?

The work here directly supports the health and maintenance of Wikidata content and the community who contributes the content.

What is, or will be, the impact of the work you did this year?

- Re-users can access and edit Wikidata's data more easily
- WMDE created a new programming interface based on the REST standard to allow technical contributors a new way to interact with Wikidata

WIKIMEDIA

ENDOWMENT

- Wikidata editors can more easily collect and maintain data about words.
- 4 more languages pass the threshold of 15,000 words. (Currently Russian, Estonian, English, Malayalam, Swedish, Latin, Hebrew, German, Basque, Slovak, Ukrainian)
- Czech has crossed the threshold in Q4 bringing the total to 19 languages with over 15,000 Lexemes.
- In total WMDE maintained relationships with many influential partners around the globe, though still primarily within the EU and NA.
- WMDE added partnerships with FactGrid, Internet Archive and Nigerian Library Institute
- 3 more organizations and/or projects outside of Europe and North America are investing resources in their own pilot projects with Wikibase
- WMDE established relationships and kickstarter projects with a diverse array of organizations interested in Wikibase from outside the EU and NA.
- In Q4, WMDE added new relationships with the following non-EU/NA partners: People's Archive of Rural India
- (PARI), Datameet, Mapbox, The Research Institute of World's Ancient Traditions Cultures and Heritage (RIWATCH), The Academy of Comparative Philosophy & Religion (ACPR)
- 520 new users and projects have successfully onboarded to the Wikibase.cloud platform

Moving forward

What is undone? The Wikidata Query Service scaling improvement work to address critical system availability and performance issues will continue into Q1 of the next fiscal year.

What are the open questions? / What don't we know yet? We don't know yet what long term technical architecture changes are necessary and sufficient to support Wikidata's continued growth and accessibility. Significant work remains to understand both the suite of technical options available to us and what changes we can decide on in consultation with the community.

Further Reading

WMDE Quarterly Reports

- Quarter 1 report July 2022 - September 2022
- Quarter 2 report October 2022 - December 2022
- Quarter 3 report January 2023 - March 2023
- Quarter 4 report April 2023 - June 2023

6 month collaboration reports

WIKIMEDIA

ENDOWMENT

- Collaboration Agreement Narrative Report Jul-Dec 2023
- Collaboration Agreement Narrative Report Jan-Jun 2023

Financial reporting

Wikidata Endowment Grant Financial Report Jul 2022 - Jun 2023

- July 2022 - September 2022
- October 2022 - December 2022
- January 2023 - March 2023
- April 2023 - June 2023