



Indonesian Wikidata User Research Report — Contributors (Editors)

Kenny Tjahjadi 7 November 2022



01 Background

Our contributors share their experiences with us.

After conducting discussion sessions with newcomers, the Wikidata Software Collaboration team held discussion sessions for contributors (1-3 years editing on Wikimedia projects) who are actively developing Wikimedia projects in their respective regions.

This research is a continuation of the entire research process to find out the response of Wikimedia contributors in Indonesia regarding Wikidata Lexeme.



Research goals

- Understand the thought process and user journey of Wikidata contributor users when searching, adding, and editing lexicographical data on Wikidata
- Explore issues that can be experienced while interacting with lexicographical data projects on Wikidata and listening to suggestions from their perspective

 Find out the current state and perceptions of contributors regarding Wikidata's lexicographic community and software development



Participant demographics

- Participants are 18+ years old
- Participants have been editing Wikidata for 1-3 years
- Participants are Indonesian speakers
- Participants belong to local Wikimedia community/have participated in Wikidata events



Research method

Semi-structured interviews, each conducted for 90-120 minutes online on Zoom. All participants turn on the camera and microphone.

Participants will be asked to do a screen share if a demonstration of the use of a feature needs to be done based on the initiative of the participant or as instructions from the moderator.

The interview is focused on exploring the problems encountered when searching for, understanding and editing lexicographical data on Wikidata.



To maintain privacy, research video recordings and participant data are anonymized.

Research hypotheses

- The contributor knows Wikidata and has heard of the Wikidata Lexemes project
- Contributors know what Wikidata Lexeme is
- Contributors can search for specific lexemes from the Wikidata site
- Contributors are comfortable with the interface of lexicographical data on Wikidata
- Contributors often edit Wikidata on computers/laptops on evenings/holidays

- Contributors work independently, without any coordination except when there are events to contribute to Wikidata
- Contributors use WhatsApp, Telegram (outside the wiki discussion feature) to hold discussion
- Contributors feel that the thing that needs to be improved the most is outreach to the community to increase the number of contributors who can be invited to edit together





02

Activities we tested

Search for a lexeme, add a new lexeme

Search for a lexeme

Current condition

Currently, we have to add *the string* "L:" or "Lexeme:" before adding a lexeme or using a tool like Ordia (<u>ordia.toolforge.org</u>).

However, this is not known to contributor users so they cannot search for a particular lexeme from the Wikidata site.

Average SEQ score: 2.8/7

Results

- 1. 4 out of 5 participants could not find the lexeme given
- 2. 3 out of 5 participants feel that there is a need for namespace filters on Wikidata search function
- 3. 2 out of 5 participants said there should be a special Wikidata lexeme search page with language filtering

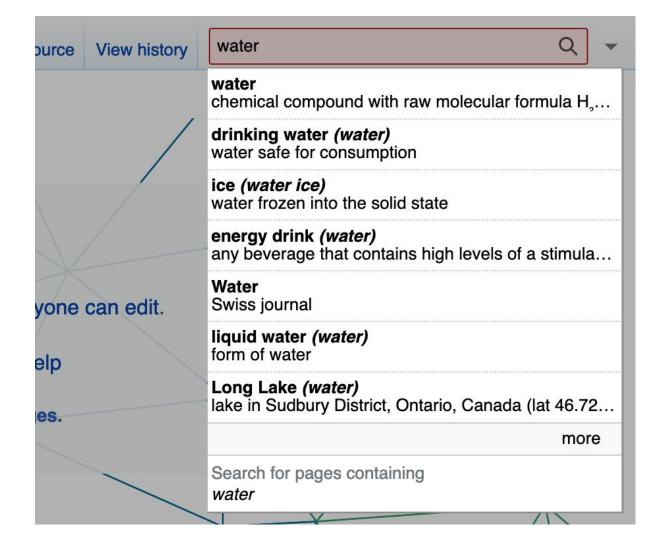


Breakdown: search for a lexeme

Activity	Search through the main page	Search through random lexemes
Journey	 User opens Wikidata site User searches by entering the lexeme directly in the search box User chooses from Wikidata's recommended results What is obtained is a Wikidata item 	 User opens a random lexeme page User searches again in the search box, expecting to only show lexeme results User chooses from Wikidata's recommended results What is obtained is still a Wikidata item 4 out of 5 users give up at this stage.
Feeling	 Frustrated because they couldn't find the lexeme, so they have to find other ways to do it 	• Frustrated because they still don't get the lexeme they're looking for
User suggestions	 Add lexeme results to search suggestions Add search filter / categorization by type and language 	 Create a special lexeme search page with search filters / categorization based on type and language

The participant could not find the given lexeme in the search results.

All search results are entries in the Wikidata item namespace.



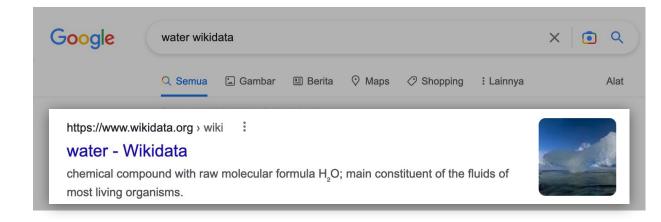


Breakdown: search for a lexeme

Activity	Search via Wikidata features	Search via external search engines
Journey	 User opens the "Recent changes" page User tries to find the requested lexeme, but cannot find it (because the requested lexeme was not recently edited) Then, user tries to use the advanced search, does not find the namespace filter so the results are still Wikidata items Performed by 3 of 5 participants.	 User took the initiative to search it via Google and Bing User sees the top search results Google displays proper results, Bing doesn't User selects the top search result Only done by 2 out of 5 participants.
Feeling	 Pressured because they can't find the lexeme even though they've explored for a long time 	• Frustrated because they have to search using Google and Bing, not from the search box
User suggestions	 Improve the filter/categorization view for advanced search to make it easier to understand 	• Improve the site's SEO make it more search engine friendly

Search results on external websites display Wikidata item results, not lexemes.

SEO for Wikidata Lexemes needs to be considered.





Add a new lexeme

Current condition

New lexeme additions are available via the side menu in the Wikidata desktop view. Contents that need to be added initially are lemma, lemma spelling variations (optional, if the script is non-Latin), language, and lexical category.

Our assumption is that contributor users can easily add new lexemes to Wikidata.

Average SEQ score: 4.3/7



Results

- 1. 5 out of 5 participants can add a new lexeme
- 2. 4 out of 5 participants felt they needed terminology guidance and examples, whether on the "add new lexeme" page, wiki articles, documents, or external videos
- 3. 1 in 5 participants chose the wrong language
- 4. 1 in 5 participants are confused looking for page links to add lexemes
- 5. 1 in 5 participants were confused about the capitalization of the lemma

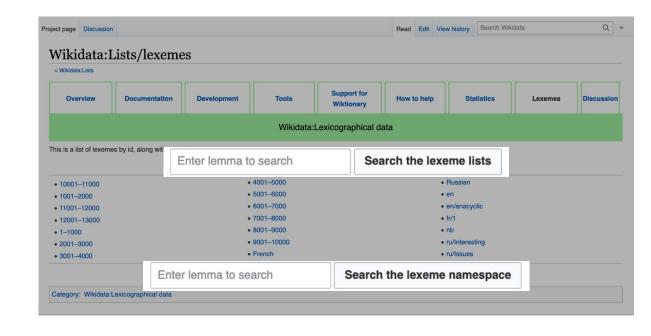
Breakdown: add a new lexeme

Activity	Open the add new lexeme page via the sidebar	Open the add new lexeme page via the lexeme information page
Journey	 User lands on the Wikidata start page User find the add link in the sidebar and open it 	 User opens the help page User selects lexeme help User skims through the Lexeme project and open the "How to help" page User opens the "Lexemes by ID" page User tries to find a lexeme via the search box by ID What is found is a lexeme with ID 1-10000
Feeling	• Glad that the link is easy to find	• Annoyed because they still can't find a way to add a lexeme even though they've used the search feature
User suggestions	Nothing here.	 Add "Add lexemes" in the "How to help" page as the main way to help with lexeme development Removed the unhelpful "Lexemes by ID" page

The Wikidata:Lists/lexemes page confuses users who want to search for duplicate lexemes before adding new ones.

There are two search boxes: the first box searches within the collected list, the second searches all lexemes.

Participants use the first search box because that is what is visible first.





Breakdown: add a new lexeme

Activity	Fill in the new lexeme data
Journey	 User fills in the lemma User searches and selects the language User searches and selects the lexeme's lexicographical category User clicks the "Submit" button
Feeling	 Confused about entry capitalization Confused by language selection having non-language results Confused with the terms used because they are only used by linguists Confused because when searching for terms, none of the suggestions show multilingual labels Annoyed at not being able to check for previously added entries Confused because they don't know how to make a lexeme with a non-Latin script
User suggestions	 Add a guide on the side of "add new lexeme" page (New lexeme page not tested) Add the option to learn about lexemes in the form of tours and articles Fix the language filter and lexical categories Add multilingual search support based on the user's pre-selected abilities, not just the display language Training is required to create lexemes with non-Latin characters



Theme	Insights
Contributor's profile	Contributing contributors are very active in their respective communities and are ready to contribute if given the right resources and incentives. Able to use Wikidata's tools proficiently. Like to learn new things and share. They want the community to thrive and be nurtured.
Learning curve	Participants need to adapt to the Wikidata view. There is nothing to search for lexemes via the Wikidata search box. Wikidata's search filters are not very good at the moment. 4 out of 5 participants had difficulty filling out the new lexeme addition form and needed help. No lexeme categorization without the need to create a query.
Community	Various communities in Indonesia have different diasporas. Some are based on region, some are based on language, some are based on Wikimedia projects that are being worked on.



Theme	Insights
Jargon	The majority of contributors are not familiar with the lexicographical jargon on Wikidata because it is too technical and the language is inconsistent. This is exacerbated by the perception of the need to learn lexicography.
Duplicate contribution	Participants want to contribute, but do not know how to check if there are duplicates in the lexeme. This causes them to be afraid of destroying the database and in the end they don't contribute.
Translation	Wikidata Lexeme's UI has not been fully translated into Indonesian or local languages. The community has not been trained to do translation.
Information sources	Almost all participants got information from Telegram and WhatsApp groups as well as social media (Instagram mostly). Banners on the main page are also effective.



Theme	Insights
Device usage	All participants disliked using the mobile site for editing. We need to create something that fits the mobile application usage pattern.
"What are lexemes for?"	All participants asked about the use of Wikidata Lexeme and its impact on society.
Language diversity	Regional languages in Indonesia are very diverse and have unique properties, such as various dialects and word levels.
Comparison with other projects	Some participants compared the appearance of Wikidata Lexeme with Wiktionary, Wikisource, and Commons.



Theme	Insights
Stockholm syndrome	All participants were not aware that the appearance of wikiprojects could be changed and improved, and felt that <i>the status quo</i> could not be overturned.
Outreach and training	All participants requested training and outreach regarding Wikidata Lexeme. They want to get material from linguists, not just from WMID. Focus on communities that can already edit Wikidata.
Contributor's suggestion	Gather new users from student clubs to leverage data, uphold the Five Pillars, protect the community from serious vandals, and remember the principle of a wiki.



Theme 1: Contributor's profile

Contributor users knew Wikimedia and its projects from social media and events held by WMID. They wanted to contribute because they want to share and contribute to society. The majority of users started contributing via Wikipedia.

Currently, they contribute a lot to Wikipedia, Wikisource, Wiktionary, and Commons projects. In fact, several contributors have participated in incubating Wikimedia projects in their local languages through the Incubator.

"I am amazed by this community. Even though it's just a volunteer project, the results amaze me."



Theme 1: Contributor profile (continued)

They admit that although WMID has been around for a long time, it is not yet visible to the public because the publication strategy still needs to be improved.

Because their motivation to edit comes from their own desires, we don't need to force them to edit. As long as the explanation is clear and their input is heard, they will immediately contribute and invite other users in their respective communities. "How come I just found out about WMID in 2021? During this time, your posts have never been recommended to me."



Theme 1: Contributor profile (continued)

The most memorable events for them are those with an element of community, such as offline meetings and co-editing sessions. They feel that this is what allows them to share knowledge and continue learning. Although competition events are also in demand, togetherness is the main driver behind it.

Because of those group sessions, they learned how to use different tools and explore features in their respective projects compared to newcomers. They have several communication channels to coordinate with the community and connect with WMID.

Community communication channels are carried out via WhatsApp and Telegram groups, while the relationship with WMID is an announcement banner on the main page and social media.



Theme 1: Contributor profile (continued)

Their impression of this Lexeme project are "dictionary" and "words", then they question the difference between it and Wiktionary. Even so, they feel motivated that the Lexeme project can become a dictionary with a unique, structured categorization, and can be used as a basis for language learning.

There are two types of contributors based on their duties in the community: There are those who actively attract new users and those who maintain a community for users who have edited. Both of these roles are equally important.

They always emphasize that they are unpaid volunteers. Therefore, they are our biggest stakeholders. They want WMID to continue to protect the community so that its members continue to grow and be maintained.



Theme 2: Learning curve

Participants said that the appearance of Wikidata is something that is not easy to understand directly. They need to take various trainings to start getting used to the interface and the content.

The ease of lexeme editing interface is considered the same as that of Wikidata items, however they admit that it takes some time to get used to the interface.

"When I first started editing Wikidata it was confusing, but now I'm used to it."



Theme 2: Learning curve (continued)

This was further proven by the absence of participants who could search for lexemes via the Wikidata search box and 4 out of 5 participants had difficulty filling out the form for adding new lexemes and needed assistance.

If users can't do these two basic things, then no one can contribute to Wikidata Lexeme, so this needs to be fixed quickly. "I give up. Can you tell me the link to the lexeme right away?"

"For new users, I think it's difficult (to add lexemes)."



Theme 2: Learning curve (continued)

Participants felt confused in editing due to the unusual interface and terms used, especially because not all Wikidata views were translated into Indonesian and regional languages.

In addition, information about the Lexeme project and in general is not on the main page or as a link in the sidebar. This makes many Wikidata users unaware of the lexicographical data it contains.

From observations during interviews, it was found that lexicographical data information pages were not helpful to users, especially lexeme list pages based on ID which were most likely created as pages for administrators.

On that page, there are two search boxes: a lexeme list search and a lexeme namespace search that the user actually searched for. However, because the position is placed below, it is not seen by the user during testing.



Theme 2: Learning curve (continued)

Wikidata's current search filter still needs improvement, as evidenced by contributor users trying to use advanced search, but unable to find filters for namespaces.

Since many contributors edit on Wikipedia as well, they expect that there is a Wikipedia-style categorization of lexemes without the need to write SPARQL query. For example, they are able to see all verbs in German.

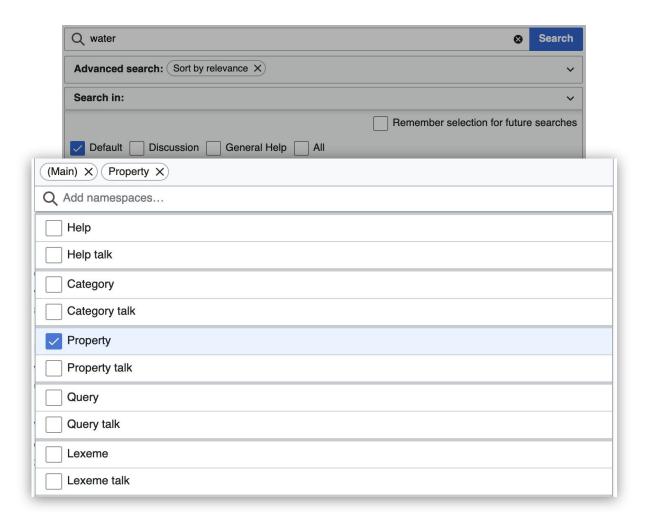
Participants suggested the following solutions:

- 1. Prepare a special tutorial for Lexeme (which is not yet on the Wikidata site)
- 2. Quick guides next to the "add new lexeme" page
- 3. Guides in the form of Wikidata articles (already exists, but hard to find)
- 4. External guides in PDF or video form



The namespace filter interface in Wikidata is not easy to understand and use.

The "Lexeme" namespace are not preselected in advanced search.





Theme 3: Community

Various communities in Indonesia have different diaspora. Some are based on region, some are based on language, some are based on Wikimedia projects that are being worked on.

Contributors in Indonesia usually follow more than one of these communities. For example, there are those who follow the language-based Banjar community and the Indonesian Wikisource community. Because of that, most community members are related to each other.

For example: when WMID announced that there was a program to document things that were difficult to do, such as *surau* (small mosques) documentation in West Sumatra, they held discussions about the representatives to be elected in the various community groups.



Theme 4: Jargon

The majority of contributors are not familiar with the lexicographical jargon on Wikidata because it is too technical and the language is inconsistent.

Terms used such as Forms, Senses, lemma, lexical category are not commonly used and must be interpreted with care and used consistently.

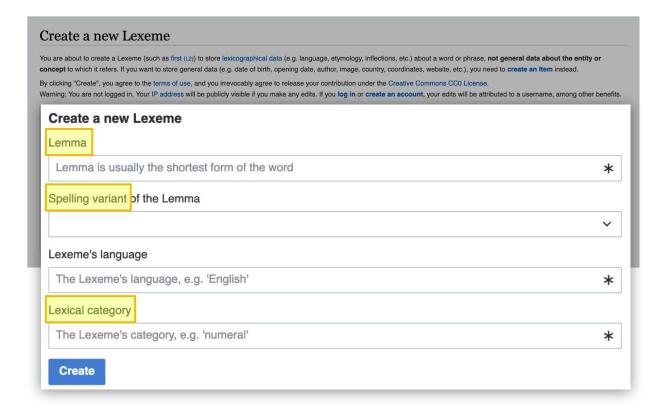
They show an example from Wiktionary where word choice is defined by the community so it's easier to understand.

This is exacerbated by the perception of the need to learn lexicography and only contributors with a lexicography background can edit lexemes in a quality manner.

"This seems to be suitable for linguists/people who like languages only."



Participants were not familiar with the lexicographic terms used, such as lemma and lexical category because they are rarely used, even in linguistics.





Theme 5: Duplicate contributions

Participants want to contribute, but do not know how to check if there are duplicates in the lexeme. This causes them to be afraid of "messing up the database" then they don't contribute in the end.

Participants want notifications about duplicate data when adding lexemes. It can take the form of search suggestions or confirmations if there are exactly the same lemma, language, and lexical categories. "If you want to add Indonesian lexemes, how do you find out if they already exist or not?"



Theme 6: Translation

Wikidata lexeme's UI has not been fully translated into Indonesian or local languages. This disturbed the editing process because the terms used were mixed with English.

So far, WMID has relied on the community to translate Wikimedia sites into Indonesian and regional languages.

Contributor users admit that so far translation has been done only by using translation tools such as Google Translate and is not double-checked extensively.

WMID has never consulted with expert translators and Badan Bahasa (Indonesia Language Development and Fostering Agency) to verify those community translations.

In addition, WMID has also never conducted localization training and made best practice guidelines because it has not been considered important.



The translation to Indonesian has not been done completely.

There are still terms that need to be discussed with the community.

air (L6546)

id

Language bahasa Indonesia Lexical category nomina

Forms

L6546-F1

air

Grammatical features tunggal, jamak

Senses

L6546-S1

Inggris water

Bahasa Indonesia senyawa dengan rumus kimia H₂O yang berwujud

> cair pada suhu dan tekanan udara ruangan dan dapat ditemukan di sungai, danau, laut, dan saat

hujan 水

Jepang

Jawa banyu

Madura aèng

Minangkabau aia

Sunda cai

Turki su



Theme 7: Information sources

Almost all participants got information from WhatsApp and Telegram groups. For information from social media, Instagram and Twitter are the most effective channels.

Surprisingly, the banner on the main page is also effective for some contributors who like to explore from one Wikimedia site to another.

Participants also suggested using email as a notification method and maximizing social media to reach new users.



Theme 8: Device usage

The majority of participants have contributed with mobile devices. However, users usually use desktop mode. Desktop mode is preferred because all the features of the page can be displayed. On mobile sites, features such as the sidebar for adding new lexemes are not available.

Apparently, contributor users remember a feature on a page by its position, not its name nor icon. When the Vector (2022) theme was applied to the Indonesian Wikipedia, they had to look again at menus that were commonly used. "It seems that mobile devices are more suitable for small edits."



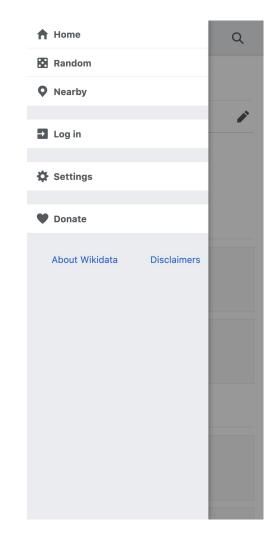
Theme 8: Device usage (continued)

Usually, the pattern of using a mobile device is light, fast, and easy to do at any time with advantages such as capturing better pictures and voice than most computers.

They suggest building an app that can ask users to do some small tasks (like adding pronunciations, pictures or references) everyday.



The mobile view does not have a sidebar with the same features as the desktop and also cannot perform full editing.







Theme 9: "What are lexemes for?"

The community is quite critical when this project is announced. All participants asked how their contribution in Lexeme will be used. On the surface, this project looks the same as Wiktionary . They don't realize that Wikidata Lexeme and Wiktionary complement each other.

In addition, they also asked about the practical application of the data. They want their contribution to benefit the public.

"After adding all the lexemes, what will it be used for? Is it just to be studied or will be used in products that can be beneficial for the public?"



Theme 10: Language diversity

Regional languages in Indonesia are very diverse and have unique properties.

For example, in Minang language, there are many dialects where people from different areas of West Sumatra pronounce things differently but the language remains the same. In addition, extra effort is needed to attract native speakers from the area who are willing to contribute.

In addition, there are languages that have a level of politeness property that Wikidata has not accommodated, such as Sundanese.

Participants are still unsure whether they can add non-Latin characters when adding new lexemes.

Wikidata Lexeme can accommodate non-Latin characters, but they need to be convinced and told how to install the tools to do so. This should be brought up when presenting training materials.

Theme 11: Comparison with other projects

Wiktionary

Wiktionary's content schema is created by the community. In addition, the explanation can be made in everyday language.

Wikistories

There was disappointment because of the different expectations regarding the platforms that can access it. Apparently, Wikistories can only be accessed on mobile devices, not on computers either.

Wikipedia

Wikipedia places more emphasis on writing skills. Participants are interested in depicting information with diagrams and tables so that articles can be interesting to read.

Wikimedia Commons

Participants are interested in adding data such as pronunciation sound files, pictures, illustrations, and videos that can be integrated with Lexemes. Videos can contain sign language as well.



Theme 11: Comparison with other projects

Wikisource

Several participants edited at Wikisource in Uji Baca program to verify the results of digitizing physical manuscript and texts with OCR.



Theme 12: Stockholm syndrome

The results of the discussion during ideation reveal that contributors tend to force themselves to adapt to the given interface rather than giving suggestions to us so that the interface is easier to use.

This may stem from previous bad precedents or the ignorance that the interface could actually be improved and customized based on community suggestions if passed on to WMID staff who could act on it.

In contrast to newcomers who give more suggestions about the interface, they prefer to give suggestions about community growth (which are equally as valid).



Theme 13: Outreach

WMID has not done much socialization of the Lexeme project, so participants who attended the interviews did not really know about this project.

The majority of participants wanted this event to be held offline, because they wanted to gather together to learn together and felt that offline events were more engaging. However, it is necessary to "test the water" by holding an online introductory event to find out the public's interest in the Lexeme project first.

Participants want to be explained about:

- 1. General knowledge about lexemes
- 2. Information about Wikidata Lexeme project
- 3. How to contribute and improve existing lexemes
- 4. The benefits of this project in the future and its relationship with Wiktionary and similar wiki projects
- 5. Training on editing, searching, and adding lexemes



Theme 13: Outreach (continued)

They suggested increasing the use of public and community social media so that many would join the event. They also want to get material from linguists, not only from WMID.

People that should be asked to contribute to Wikidata Lexeme are the ones who are able to edit Wikidata items.

Reflecting on the current Wikidata training, there was a polarization of the participants: between those who immediately understood the material and those who did not at all. This needs to be anticipated by making different events for each group of participants.

The socialization of the topics presented should be relevant and frequently discussed within the community. For example, if the audience came from an educational background, emphasize the benefits of this project for education.



Theme 14: Suggestions from contributors

Some of the contributor's suggestions are quite interesting. One of the participants suggested WMID to attract new users through informatics student associations, especially data science to process data on Wikidata for academic purposes, such as the preparation of theses and academic papers.

They want WMID to protect all of its users by <u>strictly enforcing the five pillars</u>, protecting the community from vandalism and bad actors, and to remember important things about the principles of a wiki:

- 1. Everything is done voluntarily by volunteers
- 2. Contribute; no matter how small. If there's a mistake, someone will fix it, no matter how small the mistake is
- 3. Don't be judgmental and don't be emotional



Comparison with hypotheses

- The contributor knows Wikidata and has heard of the Wikidata Lexeme project
- The contributor knows what Wikidata Lexeme is, but doesn't understand it yet
- The contributor cannot search for a particular lexeme from the Wikidata site
- Contributors are comfortable with the interface of lexicographical data on Wikidata

- Contributors edit Wikidata using computers/laptops on weekdays if there is an obligation set by the community
- Contributors work independently, without coordination except when there are community events
- Contributors hold offline meetings with the community to hold discussions
- The contributors feel the thing that needs most improvement is outreach to the community and creating a healthy, collaborative culture





04 Next steps

Next steps

Report results to WMDE

This user research can be used as a reference to understand Indonesian contributors' profile for WMDE as one of stakeholders in the Wikidata Collaboration Project.

2

Reach other demographics

This research only describes the demographics of contributors. Long-time contributors also need to be included in subsequent research.

3

Create an Indonesian contributor profile

After conducting research on all demographics, profiles of Indonesian contributors must be made by looking for common threads between participants. Differences with global contributors need to be emphasized.

4

Comparison between Wikimedia projects

There were participants who did comparisons of Wikidata Lexeme to projects like Wiktionary and Wikisource. We must find out what works and what's not from those projects as well.

Thank you