Indonesian Wikidata User Research Report — Longtimers

Kenny Tjahjadi
8 December 2022

01 Background

We talked to Indonesian longtime contributors that contributes in Wikidata.

After having discussion sessions with newcomers and editors, we talked with people who have been editing Wikidata for more than 3 years. Many of them started editing Wikidata shortly after Wikidata was released.

This research is the final stage of user research before the creation of Indonesian contributor profiles. We get many insights from them by discussing their experiences in retrospect, together.

Research goals

- Understand the state of data and Wikidata community in the early stages of Wikidata

- Understand the thought process and user journey of Wikidata longtimers 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 users' perspective

- Find out the current state and perceptions of editors' regarding Wikidata's lexicographical data community and software development

Participant demographics

- Participants are 18+ years old

- Participants have been editing Wikidata for more than 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 reflecting back on their experiences: when they started editing, getting to know Wikidata in its infancy, talking about the community and how they change year by year, and their opinions on the Lexeme project.

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

Research hypotheses

- Longtimers knows the story behind Wikidata and the Lexeme project creation

- Longtimers know what Wikidata Lexeme is

- Longtimers are comfortable with the interface of lexicographical data on Wikidata

- Longtimers edit Wikidata on computers/laptops often and regularly

- Longtimers work independently, without any coordination except when there are events to contribute to Wikidata

- Longtimers use WhatsApp, Telegram, and wiki discussion features to discuss

- Longtimers feel that the thing that needs to be done the most is making sure there's enough imported lexeme data before it can be announced to the community to make sure there are 'templates' that can be used as a reference

02

Insights gained

Insights gained

Theme	Insights
Longtimers' profile	Longtime contributors are self-motivated, love to learn and explore new topics, able to program and query, use Wikidata tools proficiently, like to edit with the community, and contribute on more than one wiki project.
Learning curve	They observed that new editors must join WMID's tutorial classes before being able to use Wikidata. No one can search for lexemes via the Wikidata search box. However, the recently updated "add new lexeme" page is easier to understand.
Social media strategy	Participants like visually appealing social media content and suggested that the content should have an educational and exploratory element to wiki tools and projects.
Internet problems	Not all community members have access to fast and lag-free internet. Make sure to create data-friendly content that doesn't weigh them down.

Insights gained

Theme	Insights
Jargon	They don't understand lexicographical jargon on Wikidata because it's too technical and have inconsistent word choice. Translation of terms need to be discussed with the community with added suggestion from linguistics experts.
Data quality	All participants felt confused when they wanted to complete an lexeme. One of the participants has created a tool that can analyze the quality of data. Inequality of data quality needs to be pointed out (a la stub articles on Wikipedia).
Duplicate contributions	The way to check for duplicate lexemes is not yet ideal. So far, this had to be done with queries and fixed with the Merge tool manually.
Information and discussion	All participants prefer to seek information through discussion channels in wiki projects, such as Project Chat page. The banner on the main page is effective for announcements. Provide discussion channels such as WikiProject for Lexeme.

Insights gained

Theme	Insights
Formal and informal language	Participants say that formal language tends to change less frequently than Wikidata items. They asked about policies for inputting informal language and phrases.
Training by longtimers	Longtime contributors should be involved in training events so they can share their tacit knowledge with other contributors. WMID can delegate the work to teach new contributors on how to contribute to wiki projects to them.
Indonesian academics and Wikidata	Not many academics in Indonesia are exposed to Wikidata and use it for research. Show them interesting studies that has been done with Wikidata data in the last 5 years to peak their interest.
Translation	Wikidata Lexeme's UI has not been translated into Indonesian or local languages fully. The community doesn't know they can do translations and haven't been trained. Not all community members understand English.

Insights gained

Theme	Insights
Query Service Improvements	If there is a syntax error in the query, an error message should be displayed. Provide a visual editor without SPARQL for easier use.
Commons usage	While some of them may use Commons frequently, they don't realize there are a lot of training materials and resources to learn from.
Reflection — Wikistories	Learning from how Wikistories was introduced to the community as a way to improve the introduction of Lexeme project.
Reflection — Incubator	Incubator development is difficult because they need to take care of plugins and has more complex markups than other wiki projects.
Ability to speak ≠ ability to write	Several languages in Indonesia, such as Minangkabau, are used more as spoken language than written language. This can impact their ability to write those words down.

Insights gained

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.
Utilization of lexicographical data	Utilizing the data on Wikidata Lexeme is the most important thing that must be done for this project to be successful. Data utilization can be co-developed with academics and the community in many areas, such as AI, NLP, and linguistics.
Editor retention	Participants questioned the effectiveness of WMID's tutorial sessions so far. Sustainability of contributors > growth. Event participants need to be followed up post-event. They believe that in-person events are more effective than online events.
Stockholm syndrome	All participants were not aware that the appearance of wiki projects could be changed and improved, and felt that the status quo could not be overturned.

Insights gained

Theme	Insight
Lexeme property bias	Participants feel that the lexeme property is too Eurocentric, since not all languages in the world have the same variety of properties.
Wishes of the participants	Participants want to be able to choose the focus of editing, track their contributions more clearly, and an interface that is more easy-to-use.
Community nowadays	So far, the editing is done individually, there is no active Wikidata Lexeme community. Participants did not understand why there is a need for differentiation between Wikidata Lexeme and Wiktionary, and its advantages.
Dialect and language	Participants consider the Wikidata Lexeme is not flexible enough to accommodate variety of scenarios for certain dialects and the ISO code is not suitable to represent regional language variations. In addition, there are concerns about characters that are not yet in Unicode.

Insights gained

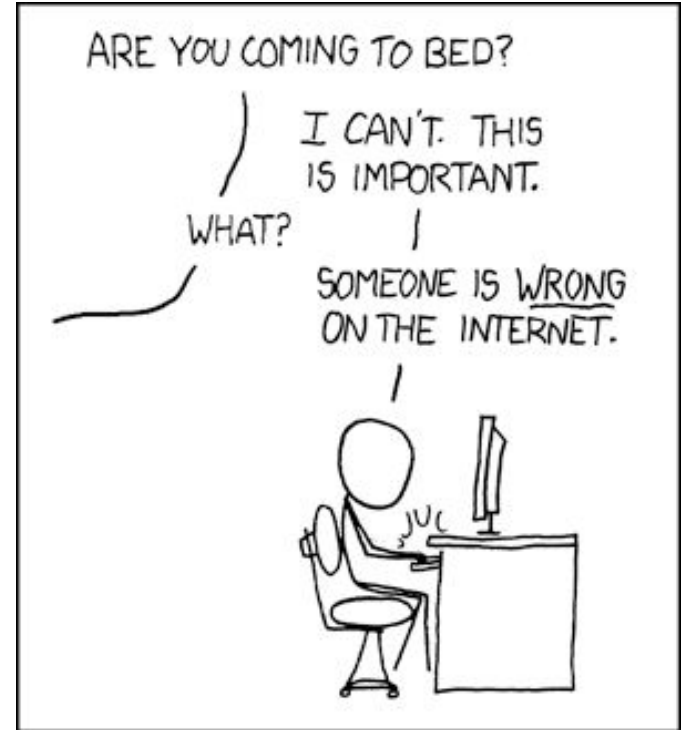| Theme	Insights
External communication	WMID needs to communicate with the community, partners and stakeholders regularly to maintain good relations.
Proactive community	The current editing incentives are not quite right. Prioritize appreciation of the community, not individuals. They want the community to be given responsibility and appreciation. It is necessary for the community to determine goals and targets together with WMID and be involved in decision-making.
Event suggestions	Give content briefs before an event starts, create more advanced tutorial sessions, show current projects that utilize Lexeme, emphasize relationship with Wiktionary.
Advice	Protect the community from vandals and political interests, reach out to communities with a sense of camaraderie and closeness, invite new contributors to join their local communities, and remember the principles of a wiki.

Theme 1: Longtimers' profile

All longtime contributors started contributing to the Wikipedia project at first. They are interested in editing because there is a personal drive that is triggered by their desire to fix something that can be changed until it is correct and complete.

Because their motivation to edit comes from themselves, we don't even need to ask them to contribute. They'll do that themselves.

Theme 1: Longtimers' profile (continued)

Currently, they mostly contribute to Wikipedia, Wikisource, Wikidata, Commons, and Incubator project. They all edit on more than one wiki project. In addition, they assume the role of administrators, contact persons, and hosts of WMID events.

They're self taught, learning through trial and error. Because they understand how to do programming and query, it's not too difficult for them to learn technical things.

"My principle when studying is ATM – amati, tiru, modifikasi (observe, imitate, modify)."

Theme 1: Longtimers' profile (continued)

They can use Wikidata tools proficiently, such as QuickStatements and Query Service. Some of them even created new tools and use Wikidata to do their academic research.

They like co-editing in-person because they feel a sense of community and feel that it can increase the retention of new editors to continue contributing.

"If we meet face-to-face, everyone will contribute together. I wish these meetings can be held again after the pandemic (COVID-19)."

Theme 1: Longtimers' profile (continued)

The topics they usually edit are topics that are close to them and their own hobbies. However, they also enjoy exploring new features, new tools, and learning new things. For example, there are participants who want to learn how to convert data from Wikispecies to Wikidata.

They love to share their learning outcomes and projects worked on their user page.

They have several communication channels to coordinate with the community and connect with WMID.

The difference is, they prefer to use discussion features in wikis such as Project Chat and talk pages over WhatsApp or Telegram, in contrast to newer contributors.

Theme 1: Longtimers' profile (continued)

Their impression of the Lexeme project is "interested", then they question the difference with Wiktionary and the urgency of making the project. Even so, they feel motivated that the Lexeme project has great potential if it can be co-developed with the community and academics.

Therefore, they encouraged WMID to seriously develop this project.

Two types of longtime contributors were found: community-based contributors and contributors from professional backgrounds (such as academics). Both are equally important and can provide insight from multiple perspectives.

They emphasize that Wikimedia projects cannot thrive without people like them who can nurture and bring togetherness to the community. They want WMID to entrust community development to figures like them.

Theme 2: Learning curve

Participants said that the appearance of Wikidata is something that is not directly easy to use, especially for new contributors. From their experience, new users need to attend various trainings to become familiar with the interface and content. Since longtime contributors are familiar with programming languages and queries, learning to use Wikidata isn't too difficult for them.

The lexeme interface is considered just as easy as Wikidata items, however they admit that it can take some time to get used to the interface.

"From my experience, new contributors must join WikiLatih (WMID's tutorial sessions) before they can use Wikidata."

Theme 2: Learning curve (continued)

This is further proven by the absence of participants who can search for lexemes via the Wikidata search box.

However, the recently updated 'add new lexeme' page by Wikidata development team is easier to understand, especially the examples given and fixed to recommendation in text fields.

The search feature needs to be fixed ASAP so that users can contribute easily.

"I can't find that lexeme. You have to use a query, right?"

"The text field recommendations when adding lexeme is already good."

Theme 3: Social media strategy

Participants appreciate WMID's visually appealing social media content such as infographics because they are interesting to read and learn about.

However, they still don't understand what WMID wants to achieve and their social media strategy going forward. So far, WMID's social media strategy is still limited to one-way announcements and still doesn't really involve the community to create original content.

They suggest that WMID's social media content should have elements of education and exploration of wiki tools and projects.

For example, they want WMID to provide infographics about new Wikidata tools that are easy to use and interesting to try, or to showcase community projects so that all Wikimedia communities in Indonesia knows what projects are being worked on by other communities.

Theme 4: Internet problems

Not all community members have access to fast and lag-free internet, especially in remote areas far from the city. There are also those who have limited internet quota due to uneven infrastructure in Indonesia. They also prefer to view content that has short durations.

This will be a challenge for WMID: provide information quickly and effectively while being data-friendly.

When creating content, make sure to keep data-friendliness in mind. For example, if you are making a video tutorial, keep it short and focus on voiceovers rather than excessive visuals to reduce file size.

Making alt text and transcripts can also help them understand the content being distributed when internet becomes spotty.

Theme 5: Jargon

Longtime contributors are not very familiar with the lexicographical jargon on Wikidata because it is too technical and have inconsistent word choice. They say that they can learn about it themselves, but say that newer users will definitely not understand those jargons.

Words used such as Forms, Senses, lemmas, lexical categories are not commonly used and must be interpreted with care and used consistently.

The solution is that translation of lexicographical terms needs to be discussed with the community in a democratic manner and with added suggestion from linguistics experts. Term choices must have clear justifications behind them and must be discussed in a centralized place.

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

Create a new Lexeme

You are about to create a Lexeme (such as first (L2)) to store lexicographical data (e.g. language, etymology, inflections, etc.) about a word or phrase, **not general data about the entity or concept** to which it refers. If you want to store general data (e.g. date of birth, opening date, author, image, country, coordinates, website, etc.), you need to **create an Item** instead.

By clicking "Create", you agree to the terms of use, and you irrevocably agree to release your contribution under the Creative Commons CC0 License.
Warning: You are not logged in. Your IP address will be publicly visible if you make any edits. If you **log in** or **create an account**, your edits will be attributed to a username, among other benefits.

Create a new Lexeme

Lemma

Lemma is usually the shortest form of the word   *

Spelling variant of the Lemma

                                                                                  ⌄

Lexeme's language

The Lexeme's language, e.g. 'English'   *

Lexical category

The Lexeme's category, e.g. 'numeral'   *

Create

Theme 6: Data quality

All participants are confused when they wanted to complete a lexeme because they don't know what statements needs to be included.

Turns out there is one participant who has felt this from a long time ago and created a tool that can analyze data quality and create entity schemas called [ProWD](). However, due to limited resources, the tool is no longer maintained.

Even though there are tools like [Recoin]() (which was made by ProWD creator's friend), it is considered to not be implemented well enough and not as complete as ProWD.

They also feel that quality imbalance needs to be highlighted to motivate people like them to complete the required statements. For example, Wikipedia stub articles have a notification banner that asks users to help complete the content. They want these kind of information to be shown on Wikidata on low-quality data.

Theme 7: Duplicate contributions

Participants can check if there are duplicates by querying and correcting them manually with the Merge tool. However, the way it is done currently is not ideal and tiresome.

Participants want to be notified about duplicate data when adding lexemes. It can be shown on search suggestions or shown if there are a lexeme with the same lemma, language, and lexical category when adding a new lexeme.

Creating a dashboard on a special page to see potential duplicate data can also be a great way to help contributors maintain data quality.

Theme 8: Information and discussion

All participants prefer to seek information through discussion channels in wiki projects, such as Project Chat and discussion pages compared to WhatsApp, Telegram, and other social media platform because they like to explore and find out information on their own.

They feel that the central notice banner on the main page is also effective for some editors who like to explore one wiki site to another.

They feel that they have nothing to discuss on global channels because none of their interests intersect. This explain the lack of Indonesian contributors in global discussions. WMID have to provide a space for them to discuss like WikiProject for Lexeme.

Theme 9: Formal and informal language

Participants argue that formal language tends to change less frequently than items, because changes are only made when there is a periodic addition of lemma to the dictionary. However, there is a kind of lexeme that continues to develop and evolve: the informal slangs.

They asked about the policy to enter slangs, for example "IDK" which is an commonly used acronym. They also asked about entering phrases such as "speed bump". In summary, what kind of words can be counted as a lexeme is still unclear to them.

This needs to be followed up by consulting institutions such as the Badan Bahasa (Indonesia Language Development and Fostering Agency) and linguists.

Theme 10: Training by longtimers

Longtime contributors argue that they should be involved in training and tutorials so they can share their experiences and tacit knowledge with other contributors. So far, training conducted by WMID has always been carried out by staff with questionable effectiveness covering limited topics.

WMID can delegate the work to teach new contributors on how to contribute to wiki projects to them. This can increase the variety of topics covered and empower communities to become active actors in the development of Wikimedia projects.

Theme 11: Indonesian academics and Wikidata

Few academics in Indonesia are exposed to Wikidata and use it for research. One of the participants is a lecturer at University of Indonesia and admits that only a few of his colleagues know about Wikidata.

Wikidata's dataset is an untapped potential for analysis and usage in academic setting, but WMID hasn't introduced academics in Indonesia to tap into Wikidata's potential for their research.

The way to attract academics to start exploring Wikidata is to show them interesting studies that has been done with Wikidata data in the last 5 years. Then, WMID needs to open a collaboration channel with them to create a research collaboration that benefits both parties. Academics get immense resources for their research, WMID gets people who can contribute and use Wikidata for good.

Theme 12: Translation

Wikidata Lexeme's UI has not been fully translated into Indonesian or local languages. This can confuse editors because the terms were mixed with English.

So far, WMID only relied on community's willingness to translate Wikimedia sites into Indonesian and regional languages.

Longtime contributors admit that when they started doing translations, they did it manually. Over time, they discussed it with the community and made general translation guidelines independently. However, language style and word choice guidelines are still based on informal consensus only.

WMID has never helped them compile procedures by inviting expert translators and Badan Bahasa (Indonesia Language Development and Fostering Agency) to verify those translations.

Theme 12: Translation (continued)

Longtime contributors use tools like TranslateWiki to do the translations. However, the way they knew about TranslateWiki because they asked WMID staff directly, so that information is known only to certain circles.

WMID should inform the community about the possibility to translate Wikimedia projects' UI so that community members who can do translations can contribute.

Longtime contributors say that translation is becoming increasingly important because not all members of the community understand English. Not all community members receive optimal English education and use it in their daily lives.

If Wikidata Lexeme's UI is not translated ASAP, it will be difficult for the community to contribute to the Wikidata Lexeme.

Theme 13: Query Service Improvements

Because longtime contributors have used Query Service for a long time, they know what improvements can be made to it.

First, the Query Service cannot return an error when there is a syntax error. The query is considered correct, but when it is executed a timeout occurs because the query cannot be processed. This is quite annoying because query syntax troubleshooting process becomes difficult if they don't even know where the error is.

They also feel that the Query Service still needs further development so that it can be more mature with feature set like other IDEs (eg: JetBrains, Visual Studio Code) by adding features such as quality of life improvements (for example, removing the need to declare `SERVICE wikibase:label`) and more extensive and accurate autocomplete suggestions like IntelliSense in Visual Studio Code.

Theme 13: Query Service Improvements (continued)

To make Query Service easier to use for people who don't understand SPARQL syntax, it was necessary to create a visual query editor without coding. Although Query Builder and Query Helper tool exists, they are not well known by the community because they have not been introduced properly. The Query Helper tool also often provides inaccurate suggestions and has many bugs, making contributors uncomfortable to use them, if they even knew about the tool in the first place.

In addition, the existing documentation for templates and autocomplete does not yet have documentation that is easy for users to understand and easy to find.

An example of a syntax error demonstration.

Wikidata only shows generic error and display the stack trace that is only useful for Wikidata developers.

Theme 14: Use of Commons

Several long time contributors often contribute to Commons by uploading various media such as photos and audio. But, they didn't know that Commons also hosts documents and training materials that had been created by WMID staff and other chapters.

They said they did not know about this because they did not expect training material to exist in Commons in the first place, because they go to Commons to contribute, not to look for materials.

All WMID needs to do is share the link to the relevant training material via project pages on various Wikimedia projects and provide them if a member of the community requests learning materials.

Theme 15: Reflection — Wikistories

The introduction of [Wikistories](#) in Indonesia is considered controversial in the eyes of the Indonesian Wikipedia community.

The community felt that they were being sidelined because WMID immediately developed the feature before listening to the community's opinion. They asked for this opinion only after the project was almost finished—a classic sunk cost fallacy that results in project release despite many objections and arguments that went unanswered.

WMID's answer regarding the reason to implement Wikistories: *"Because Indonesian people love Instagram"*, disappointed the community immensely.

This answer did not satisfy contributors because they felt that no one was going to Wikipedia to view content the way they consume Instagram content.

Theme 15: Reflection — Wikistories (continued)

After Wikistories was released, the community was also disappointed because the direction of the project was seen as unclear, there were many technical problems, it was only accessible to mobile users, the content form used only images and text (videos should have been implemented too, just like Instagram), and the need for extra moderation from vandals which burdened the work of longtimers who became administrators, which amount was seen as not balanced with total number of contributors even before the existence of this project.

Relevant discussions:

- [Original question from Indonesian contributor](#) (in Indonesian)
- [Project announcement](#) (in Indonesian)
- [Proposed implementation](#) (in Indonesian)
- [Poll](#) (in Indonesian)

Theme 16: Reflection — Incubator

Since longtime contributors are familiar with programming and tools, they tend to provide their views on Incubator development from a technical standpoint.

From that view, Incubator is more difficult than other wiki projects because they need to maintain MediaWiki templates and modules. Plus, it has more complex, non-standard markups compared to other wiki projects (since Incubator is basically a sanctioned sandbox). Contributors must edit actively and regularly, too.

"Starting an Incubator is '4×' more difficult, because there are lots of pages that needs to be completed and tools that needs to be developed."

Theme 16: Reflection — Incubator (continued)

The Wikidata Lexeme development team needs to learn two things:

1. No matter how difficult the project is, there will be people who are willing to develop them. However…
2. The Incubator learning curve results in low contributors' interest to contribute there.

This is why reducing learning curves are important for this project.

Theme 17: Ability to speak ≠ ability to write

Several languages in Indonesia, such as Minangkabau, are used more as spoken language than written language.

One participant said that this made tutorials for writing articles on the Minangkabau Wikipedia more difficult than it should be, because apparently many speakers of Minangkabau language don't know how to write those words down.

This problem intersects with the Lexeme project because the writing of lemmas must be done accurately according to the dictionary, while speech can change according to the dialect of each speaker which its addition to lexemes needs to be discussed further.

Theme 18: Device usage

All participants have contributed with mobile devices. However, users find it difficult to edit on the mobile site. The reason is users need to open multiple links in new tabs to give context to what is being edited. According to them, the snippet feature should be turned on for mobile and computer users across Wikidata as well to solve this problem.

Typically, they use their mobile devices in the morning for light administrative tasks, then switch to computers during the day.

"From a mobile device, you can add things that aren't complicated, such as just adding an 'instance of' (P31) statement."

Theme 18: Device usage (continued)

Mobile device usage patterns should be in short bursts: fast and easy to do at any time, while maximizing their advantages such as better image and sound capture.

They suggest building an app that will perform some light tasks (like adding pronunciations, pictures or references) everyday. Sensors in mobile devices such as GPS can also be an opportunity to add coordinates with precision, for example.

Theme 19: Utilization of lexicographical data

All participants asked how their contribution in Lexeme would be utilized. On the surface, this project is similar to Wiktionary and they don't know that Wikidata Lexeme and Wiktionary can complement each other.

They pointed out that utilizing the data on Wikidata Lexeme was the most important thing to do for this project to be successful.

Furthermore, they also realize that the potential for utilizing lexicographical data can bring great benefits to society.

From an academic point of view, Wikidata Lexeme can be a linguistic dataset that can be used to develop artificial intelligence systems, natural language processing, linguistic development analysis that can show relationships and derivations between languages, and many undiscovered potentials that can be developed together.

Theme 19: Utilization of lexicographical data (continued)

It will be great if utilizations of Lexeme is co-developed with academia, community, WMID, and WMDE. Cooperation and collaboration to utilize lexicographical data is the key to the success of this project.

"If developed well, the Lexeme project will produce extraordinary results! The data can be used as an interesting research material."

Theme 20: Editor retention

Participants questioned the effectiveness of the training so far. The programs created by WMID seem to only pursue the growth target of editors. They remind us that sustainability and quality are more important than growth numbers alone.

So far, WMID has the impression of releasing event participants without following up on them. Longtime contributors ask WMID to care more about them so they can continue editing and feel welcome.

"Let's honest, how big is the retention rate of WikiLatih (WMID training program)?"

Theme 20: Editor retention (continued)

According to them, in-person events are more effective than online because there is a sense of community and closeness between participants when meeting face-to-face. Informal conversations between participants are also a very effective knowledge sharing strategy that is often forgotten.

They suggest that the form of the event is a hands-on workshop because this method is more effective based on their experience in their communities.

Theme 21: Stockholm syndrome

The results of the discussion during the ideation reveal that longtimers tend to force themselves to adapt to the given interface rather than giving suggestions to us so that the interface is easier to use. They don't realize that WMID and WMDE collaboration also mean that WMID can advocate for them.

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 have more suggestions about the interface, they are more than happy to make suggestions about community growth (which are equally valid).

Theme 22: Lexeme property bias

Participants feel that the lexeme property is too European-centric, even though not all languages in the world have the same variety of properties.

This is especially evident in the Forms section, where many languages do not have it, but are one of the main elements in the data model of Wikidata Lexemes.

Theme 23: Wishes from participants

To make it easier for new users, the learning curve should be reduced by compartmentalizing the properties in the Wikidata Lexeme. This will make it easier for users who are interested in a specific field to edit only their part (e.g. IPA writing, etymology, translation).

Participants also wanted an easier view, at least similar with editing Wikidata Items.

"If the community can make better editing tools themselves, why can't Wikidata do it?"

Theme 23: Wishes from participants (continued)

Participants also wanted a place to find lexemes from the Wikidata home page, with an introduction to lexemes.

They also wanted to see a showcase of featured Lexemes whose data were considered good, equivalent to Wikipedia's featured articles.

Participants also want to have visual feedback of their contributions, or to know real world applications that use Lexeme, so that they can feel proud to have contributed to Wikidata Lexeme.

Theme 24: Current state of Wikidata Lexeme community

So far, the editing is done individually, there is currently no active Wikidata Lexeme community in Indonesia.

Participants feel that the Wikidata Lexeme project has not been the top of mind for the contributors.

Currently, there is no specific approach regarding Wikidata Lexeme to the community.

"If I'm alone in the (project chat) room for too long, I will eventually just move to other rooms."

Theme 24: The current community (continued)

The community consensus regarding Wiktionary is that its data is unstructured and not machine readable.

Although Wikidata Lexeme was created as an answer to these problems, participants did not understand the difference between Wikidata Lexeme and Wiktionary and its advantages.

The advantage of Wiktionary is its ability to provide context for a lemma in natural language and information such as etymology.

The Wiktionary community in Indonesia was concerned that Wikidata Lexeme would break their concentration and their current user base will migrate to Wikidata Lexeme, abandoning Wiktionary.

Theme 25: Dialects and languages

Participants consider the Wikidata Lexeme is not flexible enough to accommodate differences of scenarios for certain dialects.

For example, there are cases where dialects are read differently but the language remains the same, or when there is a pronunciation of a lemma that both have credible sources.

We can't rely on native speakers to do this – because they don't necessarily have the technical knowledge of the language.

Wikidata has provided ranks and multiple value feature to accommodate it, but this concept is not yet understood.

Theme 25: Dialects and languages (continued)

There were also concerns from the participant about the ISO language code used by Wikidata Lexeme.

The code is considered unable to represent the regional language variations in the field, because there are languages that have not been included, have not been recognized, and the classification is not in accordance with how it is classified in the literary books.

ISO codes for certain languages also have complex codes and are difficult to enter.

In addition, there are concerns about characters that are not yet in Unicode and how to fit them into a lexeme, for example.

Theme 26: External communication

WMID needs to communicate with the community, partners and stakeholders regularly to maintain good relations.

Communication to the community can be improved by creating content that highlights interesting projects in various communities and facilitates project discussion between communities.

They also suggested intensifying communication to partners and stakeholders and publishing a summary to the public. For example, WMID needs to reach out to academic partners in various higher education institutions and research in other areas, especially outside Java.

In addition, WMID must ensure that each staff can find out about the events carried out by WMID through an event calendar or by holding a town hall meeting (synchronization meeting).

Theme 27: Proactive community

So far, WMID only appreciates individuals, such as "best editor" or "competition winner". This causes editing incentives to be individualistic, contrasting the nature of wiki projects as a collaborative effort by many people.

According to them, community is the one who should be appreciated. It is necessary for the community to determine goals and targets together, be entrusted to develop their own programs independently to achieve those goals, and be involved in WMID policy decision-making.

"Watch out for people who claim that all Wikidata contributions are their own work."

Theme 27: Proactive community (continued)

As long as there are people who care and protect the community like these contributors, this method is considered more effective for developing a community that is mature enough.

One of the participants gave us an example: if WMID's goal is to eliminate editing restrictions due to language, the community is invited to create a translation program with their own strategy. Communities with the best translations are given appreciation after their work are evaluated by linguists and WMID.

Theme 28: Event suggestions

Participants said that our Wikidata's birthday event last October was quite good, but gave some notes for future events.

1. Provide an outline of the content before the event begins, so participants can explore the topics to be presented in advance.

2. Create advanced tutorial classes in addition to new users' tutorials.

3. Show products that utilizes Lexeme and its potential in the future, especially how the community can contribute to use the data.

4. Emphasize how Wikidata Lexeme and Wiktionary can complement each other.

Theme 29: Advice

Some of the longtime contributors' suggestions are similar to those from the editors.

They want WMID to protect all its users by strictly enforcing the five pillars, protect the community from vandals by safeguarding sensitive information, be careful of people with political interests, reach out to the community with a sense of camaraderie and closeness, invite newcomers to join their local wiki communities, and to remember the most important thing about wiki principles:

1. **Everything is done voluntarily by volunteers.**
2. **Contribute no matter how small. If there is a mistake, someone will fix it, no matter how small the mistake is.**
3. **Don't be too proud of yourself, because everyone was once a newcomer. *Stay hungry, stay foolish.***

Comparison with hypotheses

- Longtimers **knows the story behind Wikidata and the Lexeme project creation**

- Longtimers **know what Wikidata Lexeme is, but still currently observes from a distance**

- Longtimers **are comfortable with lexicographical data is presented** on Wikidata

- Longtimers **edit Wikidata on computers/laptops frequently and regularly**

- Longtimers **work independently**, without any coordination **except when there are events to contribute**

- Longtimers prefer to **explore and use wiki discussion tools like Project Chat** to hold discussions

- Longtimers feel that the things that need to be improved the most are **to empower communities, increase lexicographical data utilization, and to create a healthy and collaborative wiki culture**

03   Next steps

Next steps

1

Create Indonesian contributor profile

After conducting research on all planned demographics, Indonesian contributors' profiles can now be made by looking for common threads between participants. Differences with global contributors need to be highlighted.

2

Comparison between Wikimedia projects

There were participants who did comparisons of Wikidata Lexeme to projects like Wiktionary and Wikisource. The team must find out the strengths, weaknesses, and interesting differences.

3

Discussion of research results with WMDE

WMID needs to hold a discussion session with WMDE UX research team to compare the behavior difference of Indonesian contributors from global contributors.

4

Ideate products to be made

The UX Designer need to hold ideation sessions to find out what kind of software projects the Wikidata Software Collaboration team can develop based on Indonesian contributors' profile.

Next steps

5

Develop a UI localization plan

Wikidata Lexeme UI needs to be translated into Indonesian and targeted regional languages. This plan needs to be communicated to the community, linguists, and the Indonesia Language Development and Fostering Agency.

6

Get feedback from the community

WMID needs to get feedback from the community about product development ideas we devised from ideation **before** the product is made.

7

Collaborate with academics

WMID needs to create an introduction to Wikimedia projects that can be used as materials for collaboration, both for product development and scientific research, especially for lexicographical data.

Thank you