

Knowledge Equity in Linked Open Data

Summary of Findings

Submitted to Wikimedia Deutschland
by Abbey Ripstra at Design Research Services
May 2024

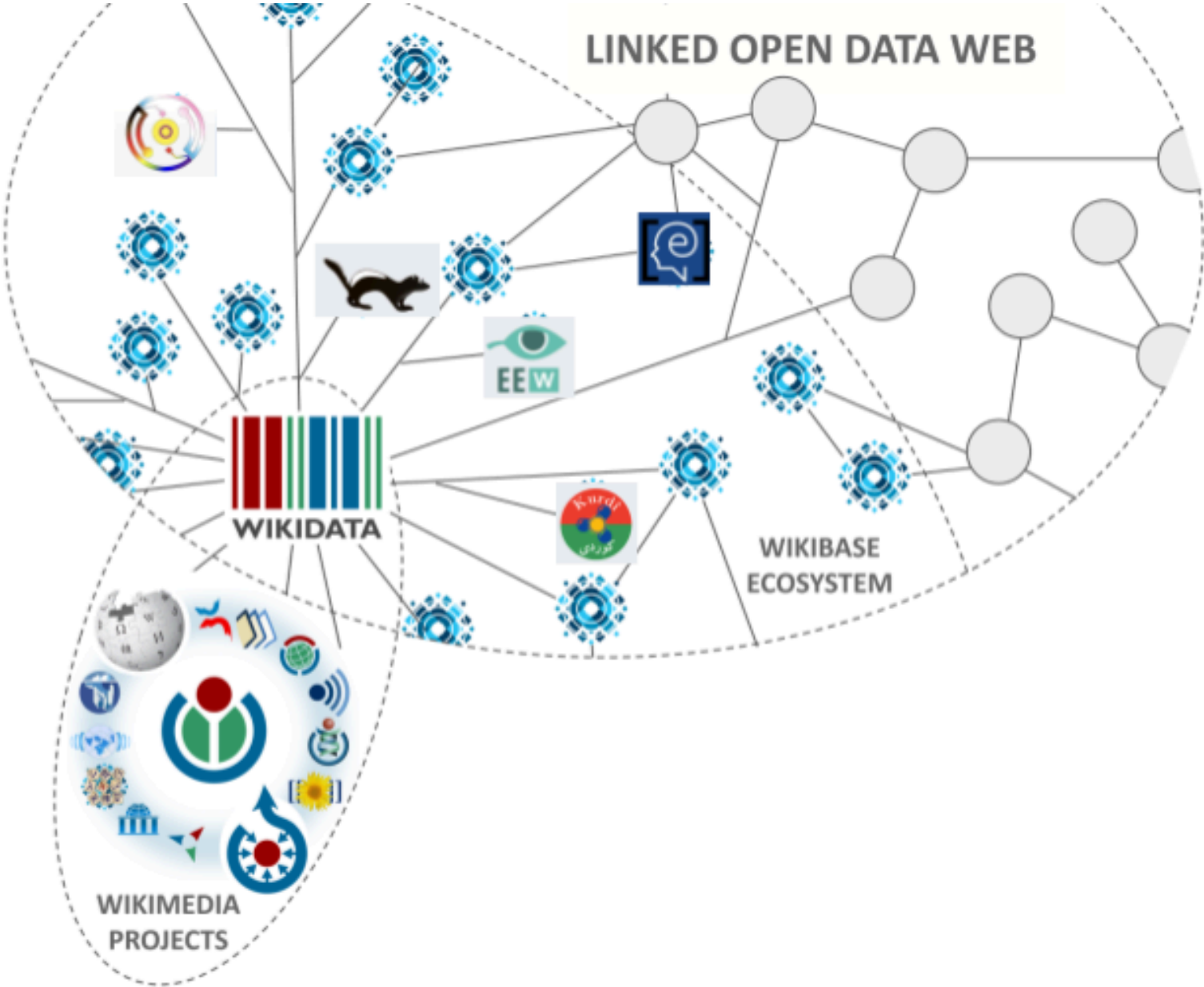


Table of Contents

[I. Introduction](#)

[II. Key Findings](#)

[Summary:](#)

[Linked Open Data \(LOD\) wide findings](#)

[LOD-wide Finding 1. Limited mobile access blocks participation](#)

[LOD-wide finding 2: LOD ecosystem providing opportunities to foster productive knowledge sharing culture](#)

[LOD-wide finding 3. Decentralized hard to discover documentation](#)

[Wikidata and knowledge equity](#)

[WD Finding 1. Need to understand concepts and possess technical skills](#)

[WD Finding 2. Wikidata's general ontology is expansive and has its limits](#)

[WD Finding 3. Difficulty working with community to accurately describe humanity](#)

[WD Finding 4: Unclear and unenforced policy about describing living people](#)

[WD Finding 5. A fuzzy definition for the notability policy](#)

[WD Finding 6. Need for consensus keeps prevailing power structures in place](#)

[WD Finding 7. Well intended editors contributing data not their own can cause errors](#)

[WD Finding 8. Turning focus from Wikidata to Wikibase Cloud or Wikibase Suite](#)

[Wikibase Cloud \(WBC\) and knowledge equity](#)

[WBC Finding 1. Resources and educational materials needed for advocates and admins](#)

[WBC Finding 2. Learning curve to understand complex concepts and specialized technical skills](#)

[WBC Finding 3. Serving as a bridge between technical and knowledge expertise, admins need more support](#)

[WBC Finding 4. Wikibase Cloud's lack of functionality makes a challenging start](#)

[WBC Finding 5. Wikibase Cloud admins need support for federation](#)

[Wikibase Suite \(WBS\) and knowledge equity](#)

[WBS Finding 1. Resource gap to self-host a Wikibase](#)

[WBS Finding 2. Learning curve to understand complex concepts and specialized technical skills](#)

[WBS Finding 3. Skills required to self express knowledge](#)

[WBS Finding 4. System functions standard in WD, missing in WBS](#)

[WBS Finding 5. Federation of Wikibases is a compelling vision](#)

[Conclusion](#)

[Appendix](#)

[For Wikidata Finding 1](#)

[Jon Kintree's suggestions](#)

I. Introduction

Project Background

Wikimedia Deutschland (WMDE) wants to foster knowledge equity in their Linked Open Data (LOD) products and as a first step, needs to better understand what the state of knowledge equity is within the LOD ecosystem; and how their products contribute to inequities within it.

Context: Wikimedia projects are powered by humans. A great deal of [research](#) has shown that projects like Wikipedia and their communities of contributors, editors, and more reflect and manifest structural and historical inequalities in opportunity and representation. At Wikimedia Deutschland (WMDE), there are early indicators that the same already is or will be true of the LOD ecosystem.

Linked Open Data (LOD): Linked open data is machine-readable, free and open, linked data that enables semantic querying.

WMDE has three products under the umbrella of Linked Open Data.

- Each in their own way, **Wikidata (WD)**, **Wikibase Cloud (WBC)** and **Wikibase Suite (WBS)** provide users with the tools to contribute to and build linked, open knowledge graphs.
- Like Wikipedia, they are all built on open-sourced software. WD, WBS and WBC are coded and designed by WMDE staff for communities who in turn contribute their data and tooling, all of which reflects and manifests social and historical structures in the knowledge it produces.

Knowledge equity: This project builds off of an emergent, internal (to WMDE) understanding that knowledge equity is a continuous and active practice. In this vision of knowledge equity, communities and data (that are part of the LOD ecosystem) represent the world in an equitable way. However, teams recognize that they lack the knowledge and skills necessary to work towards it.

In a [first exploration to identify what knowledge equity means](#), members of the Wikibase teams defined knowledge equity as the following: Every human gets to define, contribute, consume and apply knowledge in their own terms no matter what their life situation is.

In this project we seek to learn from the experience of people contributing historically and structurally marginalized knowledge using WD, WBS and WBC instances. How do WD, WBS, and WBC instances enable the contribution of historically and structurally marginalized or non dominant knowledge and perspectives? How do WMDE LOD products contribute to difficulties or barriers people face when contributing their ways of knowing? What can be changed, or improved, from a product perspective, to mitigate or address inequities and marginalization?

Note: For this project, marginalized knowledges are defined as: knowledges that have been structurally and historically pushed to the margins of society by powerful forces, whether intentionally or not, making knowledges invisible, hard to find, or inaccurately described.

This topic, learning about the experience of people contributing their knowledge and perspectives using WD, WBS and/or WBC, was chosen for this project because:

- Findings will inform product strategy and decision making as a result of:
 - a better understanding of barriers, who experiences them, in what ways, and to what extent, in their particular cultural and technical contexts.
 - identifying and sharing actionable recommendations for work that WMDE product teams can conduct to address and reduce barriers and further knowledge equity, supporting teams in their journey to understanding and furthering knowledge equity.
- There is prior work in this arena that can be built upon, and people and communities, who in contributing their knowledge and perspectives, experience barriers. Because of this, we can:
 - Invite experienced people to participate in or inform this project to learn from their prior work and take in their recommendations.
 - Connect with individuals, communities and user groups who experience this issue, and are already working to address it. This connection can create opportunities for WMDE product teams to collaborate and co-design, and perhaps even explore more equitable models for collaboration.

Project Goal

The goal of this project is to better understand how the use of WD, WBS, and WBC instances both support knowledge equity and create barriers to knowledge equity when people are contributing marginalized knowledge and perspectives. This work is being done in order to make recommendations for the WMDE product teams to address and reduce any barriers faced in this arena.

Research Objectives and Approach

For this project, research sought to better understand:

- To what extent and in what ways marginalized communities are or aren't able to contribute and represent their knowledge and perspectives in the LOD ecosystem.
- **Products as enablers or barriers:**
 - What role do WD, WBS, and WBC play in enabling the contribution of marginalized knowledge?
 - What role do WD, WBS, and WBC play in creating barriers to the contribution of marginalized knowledge?
 - What specific aspect(s) of WD, WBS, and WBC have the largest impact on people's ability or lack of ability to contribute marginalized knowledge?
 - What conclusions can be made about the relationship between WD, WBS, WBC, and the state of knowledge equity within the Wikimedia LOD ecosystem, with a focus on the contribution of marginalized knowledge?
 - Within the Wikimedia LOD ecosystem, what infrastructure or other changes can be developed to further knowledge equity?
- **Existing solutions and engagements**
 - What work is already being done with WD, WBS, and WBC to reduce any barriers and make it easier to contribute marginalized knowledge?
 - Which communities and individuals are doing this work?

Participant Selection and Recruitment

Outreach and recruitment for this project was accomplished with the generous support of the Wikimedia Deutschland Partnerships and Communications Teams, and the WD, WBS, and WBC teams who helped the researcher connect and work with the people who have the experience required to answer the questions we set out to answer. The Partnerships and Communications teams sent out notes and email invitations, along with creating a Mediawiki page describing the project and its reason for being. The WBC team designed and implemented a banner that was sent out on WBC instances inviting people to participate.

The recruitment goal was to talk with at least four people who mainly use WD, four who mainly use WBC and four who use WBS, possessing at least a novice level of experience using the products. We chose not to talk with people who are brand new and don't have any experience using the products yet.

This chart shows the participant numbers of each of the 18+ people who participated in this project, the products they use, their self reported capabilities with the products they use, and an indication of the kind of knowledge they work with is marginalized or not. Note that there were more than one participant in a few of the 18 conversations.

13 of the 18 participants work with marginalized knowledge

5 of the participants work with non marginalized knowledge *(it was helpful to compare experiences of people working with non marginalized knowledge to the experiences of people working with marginalized knowledge.)*

	Mainly using WD	Mainly using WBS	Mainly using WBC	Using a combination of the three
Advanced		P14 WD+WBS+WBC non marginalized knowledge P15 WD+WBS marginalized knowledge P18 WD+WBC marginalized knowledge	P8 WD+ WBC marginalized knowledge P16 WD +WBC marginalized knowledge	P2 WD+WBC marginalized knowledge P3 WD+WBC knowledge not marginalized P9 WD+WBS+WBC non marginalized knowledge P13 (4 people) WD+WBC+WBS both marginalized and non marginalized knowledge
Intermediate	P1 WD marginalized knowledge P17 (2 people) WD+WBS+WBC non marginalized knowledge P4 WD with marginalized knowledge	P11 WBS+ WBC marginalized knowledge	P6 WD+ WBC knowledge not marginalized P7 WD+WBC marginalized knowledge P10 WD+ WBC + WBS marginalized knowledge	P5 WD+WBC+WBS marginalized knowledge
Novice	P12 (2 people) WD marginalized knowledge			

II. Key Findings

Summary:

Linked Open Data products provide a collection of tools to accomplish equitable knowledge sharing

Wikidata is the largest free, linked open knowledge graph in the world and it attracts people to the power of WD's ability to make invisible knowledge visible and widely accessible to anyone on the internet. This power has the potential to provide the world with a wide range of knowledge from a multitude of perspectives, about many peoples, places and things. Because of this ability to share knowledge, in over 400 languages (including variants), a broad audience of people are able to take in, query and contribute knowledge using WD. Most participants learned about WBC and WBS after learning about WD.

Participants of this project describe beginning to contribute their knowledge to WD, the excitement of participating in Wikimedia projects, and being one of many people around the world participating in the Wikimedia movement. For individuals, communities or institutions holding marginalized knowledge it becomes, perhaps, even more compelling to participate in the Wikimedia movement, because it offers a way to share knowledge that has historically been marginalized, erased and even in some cases, made illegal to share. Depending on participants' understanding of the concepts required to participate and how to use the technology correctly, they faced various barriers to successfully and accurately contribute the data they hold. Specifically, marginalized knowledge was observed as being difficult in some ways to accurately describe in WD. WD, like Wikipedia, is designed from a Western, academic perspective of knowledge, and WD uses a general ontology, to describe knowledge, that is designed to be flexible enough to welcome multitudes of knowledges. Though WD's general ontology is expansive, it has bounds, and though it tries, it may never be able to model everything that humanity has to share.

Most participants described that, at some point, the barriers to contributing their knowledge accurately in WD became too great or egregious. Because of these barriers they, or the communities or institutions they work for, decide to use a WBC instance or WBS to self host a Wikibase to model and share the knowledge they hold, while not completely abandoning WD. This shift is partially because of the challenges in accurately describing the knowledge they hold using the WD ontology. Compounding these barriers is the difficulty participants described in evolving the WD ontology to more accurately describe the plurality of human knowledge. Becoming aware of the option of creating a WBC instance or of hosting their own Wikibase using WBS mostly happened after experiencing WD. The cloud and self hosting options are attractive for people who want to retain their ability to participate in the Wikimedia LOD ecosystem making their knowledge broadly visible and accessible, and to model their knowledge to be accurately described in the LOD ecosystem.

When participants described using their own WBC instance or self hosting a Wikibase using WBS, some talked about a feeling of freedom to express their knowledge in the way it needs to be expressed and described. In some cases participants described a feeling of safety and freedom of expression that comes with creating a place designed about, with, and for a community or a collection of knowledge

that has previously been marginalized. Some participants consider the vision of a future community of Wikibases, including WD, sharing a plurality of knowledges, more accurately reflecting the realities of humanity within the LOD ecosystem.

Each of the three products, WD, WBC and WBS contribute to people's ability to share the knowledge they hold, whether marginalized or not, and these products provide capabilities to make knowledge widely accessible and findable, and importantly, to co-exist with the rest of humanity's knowledges. That vision is powerful and hopeful.

This report describes, from the experience of 18+ people who are participating in the LOD ecosystem using WD, WBC and/or WBS, how each product contributes to the state of knowledge equity within the LOD ecosystem. The sections of this report describe what draws people to each product, and the challenges to, and support of knowledge equity they experience with each product. For each finding, there is an "opportunity" with considerations for addressing and reducing the barrier it describes. Additionally, there are "suggestions" for learning more about the barrier and details for how to address them, and any "requests" from the project's participants.

Though there is focus on barriers to knowledge equity in this report, this work was conducted with excitement, and optimism about the idea of greater knowledge equity in the Wikimedia LOD ecosystem, and the vision of equitably housing and sharing a vibrant plurality of human knowledges with the world.

Wikimedia Deutschland, and the Wikimedia movement at large, already have efforts underway, and resources in place which it can continue to leverage to better foster knowledge equity within the LOD ecosystem. Based on the findings identified in this project, initiatives created to further reduce barriers to knowledge equity in the LOD ecosystem should be designed with the following principles in mind:

- **Open the door to the large population of people whose main technology is a mobile phone** with little or no access to laptop or desktop computers by providing mobile capabilities for not only viewing, but also contributing to, participating in, and leveraging the power of the products of the Wikimedia LOD ecosystem.
- **Provide easily findable and accessible education to empower people with an understanding of the concepts required to contribute**, share, and query so people may confidently and effectively use their Wikibase to contribute the knowledge they hold.
- **Appeal to a diversity of people with varying experience with technology by providing up front education and documentation designed for people with varying levels of experience;** to empower people to learn to use a technology they may have never experienced before, and at the same time provide information and support for people with much more experience in making Wikibases work.
- **Describe the value of and differences between the Wikimedia LOD products to inform people about the options and opportunities** available to them. This will help people make decisions about how to best use the products for their knowledge.
- **Prioritize technical and perhaps financial support for under-resourced communities and organizations that experience time and financial barriers** to preparing and housing knowledge in a WBC instance or using WBS to self host a Wikibase, to support a broader variety of communities and organizations to participate.

- **Make support for Wikimedia LOD product advocates a priority;** including educational materials to share with their stakeholders about the value and sustainability of using WD, WBC, or WBS, along with materials to support them in training and educating people who will use the Wikibase.
- **Consider sharing future visions of how the LOD Ecosystem could work,** and the kinds of possibilities being considered, to invite people to participate.

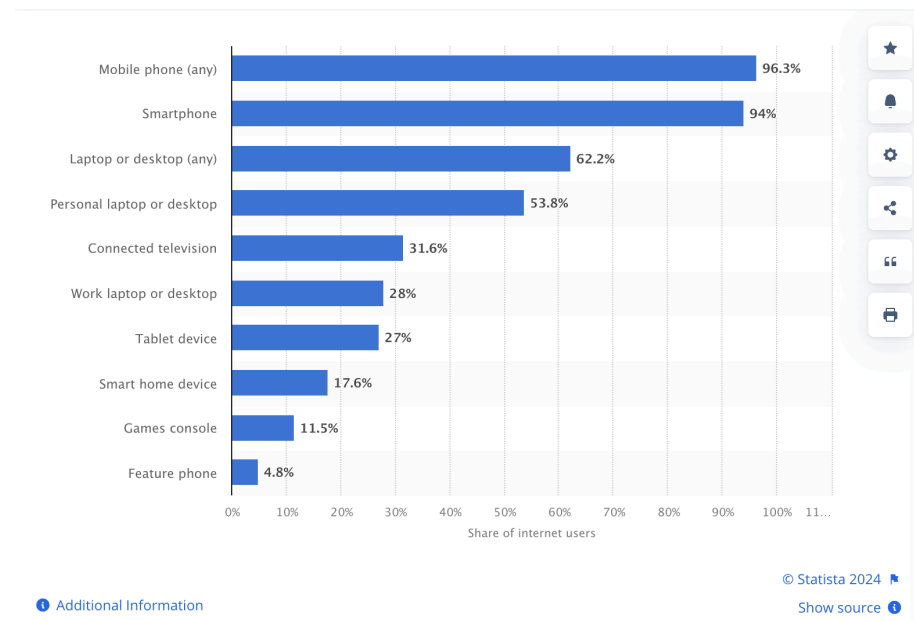
Linked Open Data (LOD) wide findings

These two LOD-wide findings impact WD, WBC and WBS users, and if addressed may serve to welcome more people to participate in the LOD ecosystem.

LOD-wide Finding 1. Limited mobile access blocks participation

From a [survey Statista conducted in mid-2023](#), it was estimated that 94.6% of global internet users use a smartphone to access the internet. In comparison, only 62% of respondents to the survey access the internet using a laptop or a desktop computer. [Mobile internet traffic accounts for about 60% of total web traffic](#), and mobile devices are how people in developing markets first connect to the internet.

Share of users worldwide accessing the internet in 4th quarter 2023, by device



For people who use a mobile device to access the internet and don't have access to a laptop or desktop computer, the lack of easy mobile access to contribute to and query WD, WBS and WBC instances, make it very difficult or impossible to participate in the LOD ecosystem. Though several participants noted a lack of mobile access as a serious issue, they were not aware of Wikimedia Deutschland teams' current efforts to make it easier to edit from mobile devices.

The [Wikidata Game](#) can be played using a mobile device and there is limited and emerging functionality for use of WD and WBC instances on mobile devices. However, this functionality is not easily

discoverable, nor is it easy to use for everyone. If one opens Wikidata.org on their mobile device, they will be taken to m.Wikidata.org, a mobile view of the product. This is the same for WBC instances. However, this view doesn't provide the ability to act. To edit statements, one needs to scroll all the way down to the bottom of the page and press the "desktop" button. Then one is taken to the desktop version of WD on the mobile device. Though it is possible to use WD and WBC instances on a mobile device, people may not be aware of it, or it may be too difficult to use. Contributing to WD or a WBC instance on a mobile device using WD's desktop version is hard because it requires zooming in and out to find things and creates inaccuracies in engaging buttons and clicking into forms along with other challenges. This all contributes to marginalized knowledge not being able to be contributed by the people who hold it. Particularly, look at India, Africa and South American mobile vs. lap/desktop usage. Providing mobile access for contribution could serve to open the door for many more people to participate.

Opportunity:

Remove barriers to participation for people who access the internet only on a mobile device to invite more people to contribute to the LOD ecosystem: Providing a mobile app or mobile web capabilities for the family of LOD products could remove a first barrier to participation and contribution to the LOD ecosystem for many people in the world. When mobile capabilities are available and ready to be released to the public, it would be powerful to conduct a campaign to build awareness about the availability of mobile access for the Wikimedia LOD products. There is an opportunity to broadcast to the broader LOD community that a mobile view exists and that more capabilities will be available in the future, with the intention of welcoming even more people in the world to participate.

Suggestions:

- Define, or conduct research to understand which features and functionalities are most important for someone to do basic mobile participation and contribution to Wikidata and Wikibases. Prioritize a small set of functionality to implement first and follow regularly with updates and improvements. The Wikidata Game and similar tools already makes it possible for people to contribute to Wikidata by acting on tasks served up from a queue consisting of existing knowledge in Wikidata.
- Connect with people who use the desktop version of WD on mobile devices, to understand which tasks they do, and how it might be better designed for mobile devices.
- Work with and learn from people who only use mobile devices, to understand their needs for mobile access to the products in their context. Internet contexts vary greatly in access, dependability and cost. One set of functionalities that works in the world of 24x7 internet may not serve everyone living in varying internet contexts.

LOD-wide finding 2: LOD ecosystem providing opportunities to foster productive knowledge sharing culture

WD, WBC instances and WBS provide opportunities for sharing marginalized knowledge accurately along with opportunities to create and foster culture to support such knowledge sharing. One reason that participants describe deciding to turn their focus to WBC or to self host a Wikibase to model and share their knowledge, rather than using WD, is the difficulty in having good faith discussions for making decisions about policy and other topics within WD.

Edit wars are not good ways to make community decisions because it is the first person to stop who loses the debate/ war. There is really no discussion or decision made, but a furious back and forth with no clear resolution. Unfortunately, edit wars happen, sometimes around controversial topics and descriptions of human beings. The culture described here does not foster productive, collaborative, good faith decision making when it happens. Participants described talk pages as places they have experienced passive aggressive and other non productive forms of communication. Even if someone has never directly had a bad experience on a talk page or been involved in an edit war, these experiences are shared and may cause a chilling effect for people wanting to participate in discussions. In a [research study](#) about discussions in WD, it was found that 30% of discussions were identified as controversial and 52% of the issues identified as controversial were related to a process.

Talk pages are and have been used for successful, productive and good faith decision making for years in WD, and many people probably have had no problems like the ones described above. Both of these experiences can be true, and if the difficulties described above happen around topics that have historically been marginalized, it may further the marginalization. The LOD ecosystem offers options that afford opportunities to foster both an ontology and a culture to describe marginalized knowledge..

Note: A few participants described difficulty in using talk pages for decision making because of the design of talk pages being difficult to use. Specifically because of having to know wikitext to successfully use a talk page, and to learn when to start a new topic and how. After sharing feedback about talk pages with Wikimedia Deutschland staff, it was shared that there has been a lot of work done on talk and [discussion pages across Wikimedia products](#). No longer does a person need to know wikitext to participate in a talk page conversation. Hopefully that it is easier to use a talk page now, and more people will be able to participate.

Opportunity:

Consider expressing that WBC and WBS provide the opportunity to foster one's own community and culture along with the opportunity to create a data model and ontology that accurately describes their knowledge. If people understand the opportunity to foster one's own culture of decision making and participating in a Wikibase, it will allow them to make a choice about participating in the culture they experience in WD or not, and still be able to share the knowledge they hold. In this way, the LOD ecosystem provides different abilities to resolve difficult conversations and decision making than, for example, Wikipedia does. The LOD ecosystem is a powerful opportunity for people to self express the knowledge they hold, as well as create a culture that fosters that expression.

Suggestions:

- Quantitatively investigate the topics that edit wars happen around and assess if edit wars contribute to the marginalization of certain kinds of knowledge being shared or not.
- Follow up on the [research](#), "The role of discussions in collaborative knowledge graphs." and investigate which topics are currently controversial. If it is still true that most of the topics are about process, look deeper to see what kinds of processes were discussed that were controversial.
- Share broadly that talk pages are updated and easier to use. Maybe not everyone knows.

- Learn from people who are using the current design of talk pages, if it has improved their experience in having productive conversations or not. Discover the aspects that have improved for people and any aspects that have not improved for people when they are participating in decision making in discussion pages.

LOD-wide finding 3. **Decentralized hard to discover documentation**

Some of the barriers participants described are compounded by a lack of easy to find documentation. Though there are multitudes of documentation for WD, it was described as widely dispersed across locations and can be difficult to find what one is looking for. For WBC and WBS, documentation is lacking or conflated with Mediawiki or WD documentation. Because of this, participants need to take more time to get answers to their questions, and sometimes need to learn by trial and error, which is costly in terms of time investment. There is a general need for documentation to be improved, more easily findable, and for there to be more documentation specifically for WBC and WBS.

Note: After sharing participants' descriptions of needing better organized documentation for WD, and more documentation in general for both WBC and WBS, with Wikimedia Deutschland staff, it was shared that WMDE is doing work to improve documentation already. All the existing resources were assembled into a redesign of the [Wikibase landing page](#). This may improve people's experience in finding existing documentation. WMDE is aware of the need for better and more accessible documentation, and are actively working on improvements.

Documentation for Wikidata (WD) is decentralized

Though there is a lot of documentation for WD, it can be challenging and time consuming to find answers to technical or other questions about how to use WD, and hard to find what one has already found. Because of this, participants described needing to learn by trial and error. This takes time, and can result in unintentionally creating more work for others and potentially causing harm of misinformation or misidentification of living people or others.

The people of the WD community are kind and helpful for people wanting to learn how to use WD, especially if one is able to find and connect to the right people or groups. The Telegram group and other arenas of support were described as positive and helpful.

Opportunity:

Learn what new and existing users of Wikidata need in documentation, and provide it in one easily accessible and searchable place. Since there is already a great amount of documentation for WD, in many forms and places, the challenge is in organizing the mass amounts of documentation to make it more easily findable and accessible by the people who need it. Include documentation for understanding important concepts required to work within WD like: linked open data, data modeling, schema, data mapping, ontologies (including specifics about the WD ontology). Also include information to answer technical questions like: how to create a WD item and the requirements and policies informing this activity, how to find the appropriate properties to appropriately describe a WD item, how to upload mass amounts of data properly, what tools and gadgets are available to make WD workflows easier.

Suggestions:

- Consider the possibility of organizing (and perhaps writing) WD documentation appropriately for people with varying levels of experience with WD. This could make documentation more approachable and usable by a variety of people with different levels of experience, and help people pinpoint the information they need more easily.
- Learn more about the inadvertent harm or errors caused by people who are just learning to use WD, and provide documentation or perhaps warnings where these errors frequently happen. Sources for this learning may be people who mentor or teach people to use WD, or by community members who patrol WD.

Documentation for Wikibase Cloud (WBC) is hard to find and much is missing

Compared to the amount of documentation about contributing to and using WD, documentation for WBC is missing, hard to find, and sometimes conflated with documentation for WD or Mediawiki. When someone clicks "help" in a WBC instance, it takes them to help about Mediawiki software in general, and not to a place to get specific support for WBC. The Wikibase User Group, the Wikibase Stakeholder Group (which is for people working in institutions, responsible for large data sets), and the Telegram channel are good places to find support from people. These groups are described by participants as the best places to learn about and receive support for WBC instances. Participants express great appreciation for the support and community engagement they get from these groups. Participants also express a lot of gratitude for free hosting and the software provided, and may hesitate to complain or reach out with requests, or repeatedly ask about fixes because of this.

"I found it frustrating trying to solve certain basic problems there regarding use of Wikibase Integrator code and datatype problems. The help on the mail list was sometimes ok but often none was given." - Participant

One participant referred to "secret knowledge" being needed to set up a WBC instance. It is clear that no one is trying to keep knowledge about how to operate a WBC instance secret, but it might feel that way. Documentation being in mostly technical spaces creates barriers for people not technically inclined. Phabricator is very useful for technical teams to organize, prioritize, communicate about and accomplish work. Participants mentioned appreciation for access to Phabricator, and after learning how to use it have no problem contributing requests and being able to see that a ticket is still in a queue or has been completed. However, not everyone who administers a WBC instance may feel comfortable using the technical structures of Phabricator. There is a lack of in app help and system feedback like error messages or helper text in WBC instances. Having in context feedback and support would save admins a lot of time and effort in finding answers to their questions. This would be especially true if in context help was designed specifically for the challenges we know admins have in preparing their WBC for being filled with data and querying well.

Opportunity:

Create and provide clear, concise, easy to find, and perhaps in context documentation, for admins to reduce the time and effort it takes to find answers to their questions for accomplishing work in a Cloud instance. Documentation should provide information on the work an admin needs to do all

along the lifecycle of a WBC instance: from the time it comes into being, to setting it up for use by people (who may not have a technical background), to maintaining the instance as it is filled with and shares knowledge.

Providing documentation in the context of the work being done, perhaps in app, or linked within the app, so people can find it when they encounter an activity they need support with, will help scale support as more people start to use WBC instances. If the questions that are known to be repeated challenges for new admins are addressed in the path of admins making their Cloud instances ready for use, there will be less need for existing support groups to provide answers to basic, easy to answer questions. This will, in turn, allow existing support groups to focus on the more challenging topics. Make sure to include education about the important concepts that are needed to be understood along with technical details for how to accomplish work in a WBC instance.

Suggestions:

- Learn from WBC admins, what challenges they had, or are having along their journey to prepare their WBC instance for use by their communities or institutions. Ask WBC admins where and how they would like to access that documentation. Perhaps some folks would benefit from visual communication, others from written documentation and others might require documentation received via audio. Once a clear understanding of which topics are most frequently needed, and how they would best receive it, design documentation so it is accessible for a broad range of people with varying technical skills to learn from.
- One participant suggested looking at what documentation there is for WD and specifically calling out, or separately providing, what one needs to know specifically for WBC instances.

Documentation for Wikibase Suite (WBS)

WBS has little documentation to guide admins through the process of setting up, preparing their Wikibase for use, and to maintain their Wikibase. Participants report support avenues being "all over the place": some on Wiki Pages, some on Youtube, in Telegram channels, and in the various user and stakeholder groups. Several participants described finding the information they were looking for once, and then having a hard time finding it when they needed it again. Written documentation for WBS seems missing and does not compare to the amount of documentation for WD. This leaves Wikibase admins searching for the answers to their questions, sometimes referring to the mass amounts of WD or Wikipedia documentation. The Wikibase User Group, Telegram channels, and the Wikibase Stakeholder Group are described by participants as positive and helpful gatherings that function as a good resource for learning and getting important questions answered.

The lack of easy to find documentation increases the time to learn, the workload for admins, and in general, the time it takes for a person to set up a self hosted Wikibase for their community or institution to use. It also potentially causes people to learn by trial and error, and potentially causing harm or more work for others or one's self.

Opportunity: In order to reduce the time and effort required to learn about and make use of Wikibase Suite, create and provide in context documentation to support admins as they work to self host a Wikibase. Creating and making documentation easy to find and access will allow people using WBS more time to focus on work that is specific to their Wikibase, like creating an ontology

appropriate to the knowledge being shared, and supporting the people who contribute knowledge to the Wikibase. When creating documentation, consider the journey an admin goes through when installing and setting up a Wikibase to populate it with data, and to welcome knowledge experts to contribute to the Wikibase. If access to the documentation is in, or linked in, the context of the work admins do, it will be faster to find. Include information on how to find and install tools and gadgets that improve workflows and make it easier to accomplish work in Wikibases that already exist in WD. Importantly, also include education on the concepts that need to be understood to create and use a Wikibase like: what linked open data is, data modeling, creating schema, data mapping, what ontologies are (including specifics about the WD ontology). Providing education on the concepts that inform using Wikibases could make it easier for admins to plan ahead and consider the future as they prepare their Wikibases.

Suggestions:

- Learn what the most frequently asked questions are for WBS admins and map them to create an in app location (or linked within the app) for people to easily find that support and documentation they need along the path they take as they use WBS. Learn about the most frequently asked questions, and what is most important for in app documentation by polling or talking with:
 - People with **no prior technical background** who have used WBS to set up and prepare a Wikibase to populate with data and for their community or team to use.
 - People **with a technical background** who have used WBS to set up and prepare a Wikibase to populate with data and for their community or team to use.

Wikidata and knowledge equity

People come to WD to share knowledge because it is free, open, multilingual, and the largest open data repository in the world. It promises visibility and access to knowledge that people, communities and institutions want to be widely available and findable by anyone on the internet. Because the LOD ecosystem is open, search engines and AI proliferate and even more widely distribute the knowledge that resides in WD and other Wikibases. Once people learn how and begin to share data into WD, they may face barriers to their contributions remaining in WD and being described accurately within WD, especially if they are sharing marginalized knowledge. The following are barriers participants in this project described experiencing.

WD Finding 1. Need to understand concepts and possess technical skills

Needing to understand complex concepts and specialized technical skills can be a barrier for people to contribute to WD because learning takes time and effort. The concepts of data modeling, ontology, linked open data, structured data, and what Wikibases are complex, and take time, effort and perseverance to learn. It also takes time and effort to learn to accomplish the technical work required to share knowledge into WD and the rest of the LOD ecosystem. Because of this, only some people, who have prior experience, the time to learn, or someone next to them to teach them, may have the confidence and perseverance to contribute to WD.

For example, several participants noted the difficulty in finding appropriate properties to describe a WD item and the complexity of workflows for knowing and finding the appropriate combination of properties to fulfill the requirements. It was also mentioned that the workflow for manually entering items into WD is

cumbersome and time consuming. It takes guess work, and running into warnings and error messages, to learn how to do it correctly with the appropriate properties.

"Browsing the extensive list of properties is challenging and daunting. Sometimes it's difficult to find the appropriate properties to model the data" - Participant

People struggle with the need to learn how to find and use the gadgets and tools that make WD work easier and more efficient. Gadgets and tools improve Wikidataists workflows, saving time and effort. Many of the gadgets are created and maintained by one person. When a gadget or tool has a bug or breaks, there can be a wait for it to be fixed. That there is one person responsible for such important tools who can be a single point of failure and at times delays or impairs people's ability to participate. Lots of appreciation and respect was shared for Magnus Manske's work, and about the availability and function of the gadgets. People also expressed concern about what happens when those tools break, and how it impacts their ability to participate.

Opportunity:

Providing inclusive, easy to understand education about the concepts required, and how to use the technology to accomplish the work could welcome a broader set of people to participate.

Education should be accessible to a broad range of people with varying learning styles, preferred languages, and levels of experience with technology, in order to reduce a barrier for people who currently may not try to contribute to WD. There are clear, easy to understand educational videos and other media that can be used to learn what data modeling is, and how to create an ontology, for people who have little experience with such concepts. Videos and other media could be produced to educate about linked open data concepts and about how to technically accomplish tasks when applying the concepts. With the inclusion of in-context help, like links directly to documentation within a workflow, or hover over information text boxes, users will have less work to find information about and learn how to accomplish tasks like adding a new set of lexemes to WD, for example. **Note:** Wikimedia Deutschland has done some work like this recently. See [Leveling up Days](#), WD [Open Online Course](#)

Suggestions:

- WD has a strong pull for people who understand how linked open data works with search engines and AI to make knowledge visible to the internet at large. Share this same concept more broadly, in a way people who don't understand these concepts could take it in. If people are aware, but don't understand how to approach, and what to do, education can create a path to welcoming participation from a more diverse group of people.
- When any warnings are presented to users, also present information for what to do to address what the warning is about.
- Participants of this project, and surely many other Wikidataists have designed and practiced well worn work-arounds which make complex technical tasks easier for them. There is an opportunity to work with people who can describe and show their work-arounds in order to inform the design of user experiences and interfaces to make complex technical tasks easier.
- Participants pointed to Open Street Maps as an example of a functional way to make it easier and less time consuming to correctly contribute data to WD via a form:

- "For example, if we want to add an item about a book, we could simply search for "book," and then a wizard would be shown along with an easy-to-use form to add detailed information about that book." Participant
- Mentioned by at least three participants, was an idea for using natural language processing to create forms to more easily and properly define data being contributed to WD.
 - "Perhaps we should consider developing a user-friendly scripting language that allows users to add a collection of interrelated items all at once, instead of creating and modifying each item individually using a PC-based web browser." - Participant
- One participant, to make finding the correct properties easier, found [a table containing over 12000 WD properties](#). After finding this table, the participant was able to manually browse the table to find the properties they needed and reorganized them using a thematic organization (see attached screenshot in the appendix) for easier future reference, reducing their work when describing items in WD.

Requests:

- By providing a sample of a WD item correctly described and referenced for an example area of knowledge, people could see how to model correctly and learn through that example. Perhaps there could be several examples. **Note:** There are examples like this in some places, but they are not visible, or findable in all the places they are needed.
- Though staff from Wikimedia Foundation and Wikimedia Deutschland support people working in GLAM institutions and using WD, a request was made to also have support from staff who are technologists to answer technical questions for people working in GLAM institutions. **Note:** In response to this request, it was called out that the WMDE Partnerships team does outreach and answers questions for people using WMDE products in institutions. Also, for WD, more technical questions are answered on-wiki by developers, product and community communications staff.
- Requests were made for more support from Wikimedian technologists who also understand linguistics and how language is structured for humans and for machines to take in.

WD Finding 2. Wikidata's general ontology is expansive and has its limits

Wikidata's general ontology having limits for what it can describe accurately, motivates people to turn their focus from WD to share knowledge in a Cloud instance or to self host a Wikibase using WBS. WD and Wikipedia are designed and structured based on a western academic epistemic frame. A Western academic epistemic frame is one way of considering and understanding knowledge - through Western academic perspectives. [Epistemology](#) is the theory of knowledge, how we understand what knowledge is, where it comes from, and what is true. There are many kinds of human knowledges, and a western academic knowledge is just one kind of knowledge.

[Ontology](#) is a way of describing the properties of a subject area, or kind of knowledge, and how the properties are related. An ontology is created by defining a related set of terms and expressions that represent the subject area - or knowledge accurately. Though the intention is to have the WD ontology be general enough to accept many kinds of knowledge, the general ontology has bounds. Functionally, the ontology is a consensus ontology, because decisions about change require consensus of people in a discussion to modify the ontology. The consensus ontology of WD is not necessarily a good fit for knowledge that has been marginalized, or for knowledge that is structured or shaped differently than a Western academic form of knowledge.

“The Wikidata model doesn’t fit the model of the history we are trying to tell” - Participant

“Wikidata, as much as it tries will never be fitting for everything” - Participant

Perhaps because of this, there are undefined or incorrectly defined items in WD. The general ontology of WD can keep some forms of knowledge out of WD, and can make it so knowledge is inaccurately described within WD. For example, WD only has Gregorian and Julian calendars for use. Some communities have a different concept of time that is very hard to map to the WD ontology. The way sex and gender identity are described in Wikidata, with the conflation of sex and gender in one property, functionally marginalizes accurate knowledge about LGBTQ+ and Queer communities and peoples.

*“Some of the constraints on Wikidata properties are either overly verbose -- requiring users to add basic ‘common sense’ facts that could be inferred with sufficient context -- or overly restrictive - only selecting highly specific definitions while disregarding several possible interpretations”
-Participant*

When incorrect properties are added to an item queries can be inaccurate as a result. Participants report finding content unrelated to their query in results, or items that they know to be in WD but missing from a query. This structure further marginalizes knowledge that is not able to be accurately described in WD.

Opportunity:

Consider sharing information to create greater awareness about epistemology and that there are many kinds of human knowledges existing together, at the same time in humanity. Include information about what ontology is, and that there are opportunities to create an ontology appropriate for the knowledge being shared (using WBC or WBS). Acknowledge the WD general ontology as it is, with its expansiveness, and its limitations. Creating greater awareness of the concepts of epistemology and ontology, and about different options for modeling knowledge, could reduce the amount of time and effort people spend when their knowledge may not be best modeled, or described by WD’s ontology.

Suggestions:

- Research how the concepts required to understand linked open data are shared in easy to understand ways on the internet to inform how to clearly share these complex contexts with a range of people who have varying backgrounds and varying understanding of the concepts.

WD Finding 3. Difficulty working with community to accurately describe humanity

Difficulty working with the community to more accurately describe humanity motivates participants to turn their focus from WD to share knowledge in a Cloud instance or to self host a Wikibase with WBS. When attempting to contribute marginalized knowledge, sometimes people face further marginalization when suggesting changes to the ontology, proposing new properties or

contributing knowledge that can not be accurately represented within WD's ontology. It can take a lot of time to successfully propose and accomplish a new property in WD.

“ I found Wikidata to be limiting due to the need to propose new properties and wait for a committee to approve them. So I switched to my own Wikibase.cloud project.” - Participant

In an attempt to honor the work of the WD community and WD's general ontology, and at the same time accurately represent their knowledge, people approach the WD community in hopes of adjusting the ontology and properties. Beyond accuracy, this is important work for protecting people from harm as a result of misrepresentation, or by describing living people without consent which can put people in danger in the physical world. Misinformation and mis-identification, outing, deadnaming, and other harms can spread quickly and multiply when proliferated across the LOD ecosystem and the broader internet.

The experience of working with the WD community to propose new properties, adjust or correct existing properties, and introduce policies about how to describe living people to reduce harms, has been described as frustrating, exhausting, a political project every time, not possible, and even harmful. Particularly for people contributing marginalized knowledge this experience may feel like further marginalization.

Opportunity:

Support people, new to WD, to understand broadly, what data modeling is and ontologies are, and specifically what WD's general, or consensus ontology is. Provide information to create awareness that there are options for differently modeling knowledge by using WBC and WBS.

Sharing a view into the options of WBC and WBS for modeling knowledge and describing it accurately with ontology, may be encouraging to people who have tried and failed to adjust the general ontology to accurately describe the knowledge they hold. This information could alert people to consider the possibilities of using WBC or self hosting a Wikibase using WBS. Providing this information, as people start to learn about and work in WD, may help them understand the limitations of a general ontology, and support their planning.

WD Finding 4: Unclear and unenforced policy about describing living people

Unclear and unenforced policy about describing living people / Property 21, and difficulty making change, motivates participants to turn focus from WD to share knowledge in WBC or to self host a Wikibase using WBS. Difficulty with the use of [Property 21](#) (a property conflating sex and gender) and lack of [clear policy and enforcement](#) of policy describing living people were described by almost all of the participants. The attempts to address these concerns end up in unresolved conversations and no solution for a long time. Lack of resolution of issues like this results in the potential of proliferation of incorrect information and mis-identification of people, and harms. For example, harms like dead naming, misgendering and outing people can occur. It is not safe for people to be identified in a certain way, and there is no consent process for biography or description of living people. If a LGBTQ+ or Queer person is

living somewhere where it is illegal to be who they are, being outed in the LOD ecosystem can be very dangerous.

“There have been and continue to be many attempts to address the issues with Property 21 [sex and gender]. The minoritized group has strong feelings about it with good reason, but the decision making process around it is made by the majority so the people most affected by decisions, or lack of decision making, are overridden” - Participant

Most of the participants, after working to change or update a property or focus on policy for how to describe living people, at some point, decided to move on and not attempt to change it. Only a few participants continue in the effort to evolve the WD ontology as humanity evolves, while also sharing knowledge on another Wikibase (either self hosted or a Cloud instance).

Opportunity:

Work to clearly define and create structures to enforce the WD policy on describing living people.

Ensure that when reversion or deletion happens, that there are clear reasons shared to encourage learning and understanding of the policy. Consider adding a consent process to the policy on describing living people so people can provide or decline consent for descriptors to be added to their WD item.

Suggestions:

- Exchange ideas with the Wikimedia Foundation legal and support teams and the Arbitration Committee of English Wikipedia to take inspiration for how the policy on biography of living people could better protect living people from harm when they are being described in WD.
- Noted by a participant as a potential method to protect individuals from misgendering: It would be nice if you could put a block on a property that is describing you as an individual. Even better if a block could stick and if it could erase misgendering from the history of the item. **Note:** WMDE staff informed that individual revisions/edits can be hidden from the history of an item by admins on request.

WD Finding 5. A fuzzy definition for the notability policy

A fuzzy definition for the Notability Policy can lead to deletion of knowledge and data without clear reasoning, motivating people to turn their focus from WD to share knowledge in a Cloud instance or to self host a Wikibase using WBS. Contributing to further marginalization of knowledge, is the vague definition and quick application of the notability principle, sometimes with a lack of a clear explanation. For example the [WD notability general guidelines](#) describe how an item is referenced acceptably to remain in WD as : *“It refers to an instance of a clearly identifiable conceptual or material entity that can be described using serious and publicly available references.”* Because of the vagueness of “serious”, there is potential, and room for deletion with little reason. This manifests in content being shared by Wikidataists and then quickly being deleted because it was deemed not notable enough, or not referenced according to the standards of WD.

“We find that the Wikidata notability criteria really serve to prop up

existing power structures, in ways that we don't particularly like. For music specifically, they would prefer that only musicians who have released with a record label or published under a publishing house be added to Wikidata." - Participant

This results in a lack of learning because contributors don't always understand the reason for their contributions being deleted. It also results in a lack of ability to share knowledge, and potentially a chilling of the desire to share knowledge.

"Notability is a tool for exclusion. In Wikipedia and Wikidata, one person can make a deletion or merge... though they don't necessarily know the data." - Participant

This experience can feel like a strong rebuff for those who invested time and energy to find references and research content to be shared into WD. For people who are attempting to share marginalized knowledge, it can take more time to research and find references in support of that knowledge - than it may take for knowledge that is not marginalized. Why does some knowledge need more references than other knowledge? Sometimes content is deleted without any clear reason.

It can be more work for people to successfully find references and cite marginalized knowledge in WD and Wikipedia, than for people sharing knowledge that is accepted in or is part of a Western academic epistemic frame. Marginalized knowledge has already been erased and made invisible by prevailing power structures in the physical world, so those looking to share it need to do more work and spend more time to defend and assert the validity of their knowledge than people sharing knowledge that is not marginalized.

Opportunity: Support experienced Wikidatists who patrol contributions to WD to provide reasoning for any deletion that occurs. Providing this reasoning could create a friendlier welcome, an easier path for new Wikidatists to learn what is notable and what is not notable in WD. With information about why something is deleted or deemed not notable, there is an opportunity for discussion and better understanding about WD notability policy, and how to apply it. Perhaps there is room to revisit the notability policy to adjust the strictness or openness. There are benefits to a clear and strict policy in that it is understood by all where the bounds are. However, those bounds could serve to marginalize some kinds of knowledge depending on the definition of notability. One benefit to a more open notability policy, accepting a broader range of references, is that more kinds of knowledge may be contributed and survive in WD. Having a clear definition, and perhaps Wikidataists revisiting what kinds of references are acceptable and not acceptable under the Notability policy, may serve to welcome a broader set of knowledges to WD.

Suggestions:

- Ensure that new Wikidataists, just learning the policies and practices, are aware of and have information to understand the WD notability policy so they aren't surprised if they encounter their contribution being deleted as not notable enough.

- Share with new Wikidataists that if their content is not notable enough for WD, they can create their own notability policies in their own WBC instance, or use WBS to self host their Wikibase in order to share the knowledge not acceptable in WD. People may not be aware of WBC, or WBS, even though they know about WD. **Note:** WMDE added a note about Suite and Cloud to the Notability policy page for this very reason.

WD Finding 6. **Need for consensus keeps prevailing power structures in place**

The need for consensus keeps prevailing power structures in place, and motivates people to turn their focus from WD to share knowledge in a Cloud instance or to self host a Wikibase with WBS.

Because consensus is required for decision making in the WD community, prevailing power structures remain in place. Sometimes when a large majority of participants in a discussion agree to make a change, one person can disagree, and then the decision is never made.

"The idea of consensus is unclear. Once having disagreement - and not finding complete consensus - there is no referee - it is space for a bully to win." - Participant

The issue of Property 21 is a prime example of this. For years, there have been, and continue to be, numerous discussions about and proposals to change Property 21 to more accurately describe people in WD. No change has been able to be made. A complicating factor about Property 21 is that it is required for describing an "instance of" a human being. A warning error is thrown if a WD item is being created and this identifier is not added. This furthers the problem by making it required to, in some cases, misidentify an instance of a human being contributed to WD, because the person contributing the property does not know the status of the person being described and guesses. The status quo can remain in place with a lack of ability for equitable decision making about properties like Property 21.

Opportunity:

Continue to explore alternative, and more effective ways of making decisions in a reasonable amount of time in WD, so decisions are not held off by the need to attain consensus. There are many examples of successful decision making processes and strategies that could be used in WD to come to decisions about controversial issues that can currently take a long time to resolve.

Note that in a 2023 paper, titled "[Agreeing and Disagreeing in Collaborative Knowledge Graph Construction: An Analysis of Wikidata](#)", the authors conclusion includes

*"Our main findings are that quick decisions are taken in the creation of properties, the majority of controversies are related to policies and practices, and that **more than half of controversial discussions do not lead to consensus**. Therefore, we recommend to the Wikidata community to pay attention to the quick decisions regarding properties which still need to be explored in order to understand the impact on the quality; enable clear and easy access of policies and rules to help editors participate efficiently in argumentation; **and improve strategies for making consensus in long and controversial discussion so as to avoid unresolved issues.**"*

Suggestions:

- One participant suggests a possible solution: In activist groups sometimes if there is an issue, a subgroup is nominated and voted in that is then empowered to make a decision about the issue, and address the issue. This group is disbanded once the issue is addressed and solved.

Perhaps this method could be used to address particularly troublesome problems like the negative impacts of Property 21, conflating sex and gender, in describing human beings.

- Another suggestion from a participant to address this issue, and intractable issues like it, is to fund an action committee of people to hear issues and feedback about issues in WD from the community. This committee would be resourced and empowered to make decisions to fix problems.

WD Finding 7. **Well intended editors contributing data not their own can cause errors**

Well intended editors sharing knowledge that is not their own can cause errors and potentially harm people. Sometimes people learn how to use WD, and with good intentions want to contribute a lot of knowledge to WD and the LOD ecosystem. It is beneficial to the LOD ecosystem to be able to easily contribute mass amounts of data. However, there can be drawbacks. Because it can be easy, once one learns, sharing mass amounts of data into WD can create issues like mass misgendering, or other descriptive inaccuracies and harms. One participant described noticing bot edits doing mass misgendering, probably unintentionally, on WD. The participant figured out how to use a bot to correct the inaccuracies. This took a lot of time and effort on their part, and the issue of mass misgendering still is observed from time to time. Not everyone would know how to accomplish that correction. When this happens, it is another form of active harm and marginalization, even if it is unintentional. One participant mentioned that because of the potential for harm caused with mass contribution (especially when describing human beings), Wikidataists could slow down and to take time to consider the possible consequences, of even small margins of error. Even a very small margin of error, when data concerns living people, could cause harm to a person. Taking time and careful consideration when making mass contributions about human beings could contribute to avoiding harms caused by the good faith actions of Wikidataists, and may reduce the need of other Wikidataists to fix errors as described above.

“Creating biographical items was quite easy because Wikidata offers a form and requests entry of the next item. Once you enter that an item is human, then Wikidata suggests what to add next. This guidance for adding properties is useful and makes it faster and easier to add properties. Because of this, it’s not always easy to know what you are doing. There is no guidance for how to use properties accurately within this workflow, so it is easy to miss-identify people.” - Participant

This happens not only in mass contributions of data, but with individual items and properties. In one case, a participant’s home village is an item on WD. When another editor changed the spelling of that village, it was then an inaccurate spelling of the village in WD. The participant, who grew up in that village, had to have multiple conversations to have the spelling of the town they are from spelled accurately on WD. After some difficult conversations, the issue was resolved and the village is now spelled correctly.

It can be difficult to change inaccuracies in WD, and to ensure that a correction will remain in place. The kind of engagement described above, when trying to correct information one knows is inaccurate, can

cause more work and time for people who already experience marginalization in both the digital and physical worlds.

Opportunities:

Consider designing workflows to slow Wikidataists down, especially when contributing data about living human beings, to consider the potential for and avoid inaccuracies and harm. Learning or understanding how most misgendering and other harms happen within WD workflows could inform which workflows to pay further attention to, and redesign to support careful consideration. By slowing people down and providing information about potential inaccuracies and harm that could be caused by their action, and how to avoid inaccuracies or harms, it may be possible to reduce those inaccuracies and harm. Especially in an environment where people need to learn a lot to participate successfully, there may be an over emphasis on designing a user experience to be easy or fast, when consideration is ultimately more important than speed. Making it easier to accomplish certain tasks quickly is powerful, but when those tasks or workflows could potentially cause harm, there is an opportunity to slow the experience down and create space and time for consideration.

Consider how to better support Wikidataists who identify misinformation or misidentification of people or items in WD, and want to fix it. Are there successful methods not generally known that could be more broadly shared? Is there education that could be provided on how to use bots or gadgets to address misinformation or misidentification? Work with WD patrollers, and other Wikidataists concerned with correcting inaccurate data, to learn about their experiences in noticing and addressing misinformation and misidentification when they encounter it. Use this learning to design processes, and workflows for kindly and accurately correcting misinformation and misidentification in WD.

Suggestions:

- Learn where and how mass misgendering and any other harms or mass misinformation happens. Focus on solutions for those specific workflows and supply education for Wikidataists about the risks, and what they can do to mitigate the risks as they work within WD. In any warnings or methods for slowing contributors down for consideration, use clear and concise language for describing both potential risks and methods for mitigating the risks. Avoid dramatic or frightening language or design so people are willing to take in the information, and at the same time, design to attract people's attention to the topic for consideration.

WD Finding 8. Turning focus from Wikidata to Wikibase Cloud or Wikibase Suite

Participants described turning focus from WD to WBC or WBS because of their need to model knowledge differently than it can be modeled in WD. Participants described first using WD to contribute to the Linked Open Data ecosystem, and then adding either a WBC instance or self hosting a Wikibase using WBS. This decision was informed by and made after experiencing difficulty in making decisions with the WD community when proposing changes to properties, and proposing adjustments to the policy around describing living people. Also contributing to the decision, was the need for language or country tags that were not granted, and other issues that go unresolved because of the inability to effectively and fairly make decisions about contributions and WD policies.

[And the community is changing because of the lack of being able to be in

the sum of human knowledge as it is - this is because the community has had to shift to fit into the prevailing structures in society in general, and in Wikidata. Traditionally, our language uses gestures and oral communication, not written language] - Paraphrase of a Participant's point

A few of the participants had no problem working with the WD community because they never needed to change a property and their data consists of academic research papers that are not necessarily marginalized knowledge. These participants didn't mention having issues with the descriptions of the knowledge they contributed. Another small group of participants, though they experience the issues described above, for various reasons are continuing to work to make change within WD to more accurately and equitably describe knowledge.

Most participants decided at some point to add the use of a WBC instance or self hosting a Wikibase using WBS to house their community or institution's knowledge rather than focusing solely on contributing to WD. Several participants described the contrast between the constraint of librarians using authority files (detailed descriptions, collaboratively defined, of how to describe a particular collection or kind of knowledge) and the lack of clear definitions of properties in WD. This issue causes pause and concern for people who are responsible for large sets of data that they work with in libraries, museums, government or academic institutions that want to participate in the Wikimedia LOD ecosystem via WD. That optional avenues exist for modeling one's knowledge in the way it needs to be described to be accurate by using WBC or WBS is a valuable opportunity for those experiencing difficulty in contributing to WD.

Opportunity:

Leverage the powerful draw of Wikidata making knowledge widely visible and findable, to create more awareness about WBC and WBS, and about the collective power of the Wikimedia LOD ecosystem to contribute multiple forms of knowledge. WD may be the most known of the three products, and participants described being drawn to WD's power to make the invisible visible, and to support them in their effort to make the knowledge they hold widely discoverable and usable. WBC and WBS provide, in concert with WD, an even broader opportunity to model, share, and make visible a plurality of knowledge. Creating greater awareness that WBC and WBS can be used to model and share multiple kinds of knowledge, alongside and in concert with WD, could serve to reduce marginalization of knowledge in the LOD ecosystem. The vision of multiple kinds of knowledge existing alongside each other could be a draw for many who hold marginalized or otherwise sidelined knowledge. The challenge here is in communicating this information in a way that people, even with little experience with the LOD ecosystem and its products, can understand the value of.

Suggestions:

- When people start to participate in WD, communicate that WBC and WBS exist, and what each product offers, including how they differ from each other. Importantly, also communicate about the value of, and how the three products can work together.
- Create and provide a one page product comparison chart on clearly describing the differences, and similarities, and how the products work together.

Wikibase Cloud (WBC) and knowledge equity

Participants describe being attracted to WBC's free hosting and the ability to model and describe knowledge the way it needs to be described for their project and for accuracy. Along with this appreciation, participants described barriers they experienced in using a WBC instance to share the knowledge they hold, which are shared below.

WBC Finding 1. **Resources and educational materials needed for advocates and admins**

Providing more resources and educational materials for Wikimedia advocates could make communicating the value of WBC instances for sharing knowledge, and how to use it, easier for advocates and admins. For some institutions and communities, there may be hesitation to make the move to a WBC instance. This hesitation is fostered by questions like: How is this service funded? Will it be sustainable into the future or will it disappear some day? Will it stop being free at some point? Is the data safe in a WBC instance? One participant reported proposing the use of a WBC instance to house governmental data. Though the participant is knowledgeable about the value and structures of WBC instances, they were not able to convince the stakeholders, who would potentially approve and fund such a project, of the value of WBC and the need for using it versus WD or another open data repository. People who advocate for and support organizations and communities to use WBC instances not only need to convince people about the value of the product, they also need to educate and train people about the concepts required and how to contribute knowledge and query and share the knowledge in a WBC instance. This all takes a lot of time and effort, and anything to save time and effort would be helpful for admins who are also advocates. For smaller, or less well resourced organizations and communities, having the resources to learn to use, set up and facilitate a community or organization to use a WBC instance may be out of reach, or a stretch.

Opportunity:

Provide support for WBC advocates so they can more easily communicate the value of using a WBC instance to both large institutions and small organizations. Additionally, provide educational materials advocates can use for teaching people how to use a WBC instance. Providing this kind of support could save advocates time and effort as they work to advocate for the use of a WBC instance, and as they prepare people to use one. Learn more about what questions and concerns advocates encounter as they work with people who are considering using a WBC instance, and use that learning to inform what additional kinds of support advocates may need. Learning what specific information has been helpful for people who successfully advocate for an institution or community to use a WBC instance could also be useful in focusing support.

WBC Finding 2. **Learning curve to understand complex concepts and specialized technical skills**

The need to understand complex concepts and specialized technical skills to populate WBC instances, can be challenging, and slow admins down when preparing their instance for being populated with knowledge. To successfully populate and use a WBC instance, people need to understand the concepts required like: linked open data, data modeling, what an ontology is and how

one works, for example. Not only do people need to understand the concepts, but they need to know how to apply the concepts, using the technology and tools provided in a WBC instance to model data, create an ontology reflecting their knowledge so their instance can properly ingest and share the knowledge they hold, and to query their knowledge. If people are not experienced with the technology and how to use it for data modeling and creating ontologies, it can be a blocker to getting started with a WBC instance. Participants described experiencing a lack of knowing what to do when confronted with the blank canvas of a new WBC instance. WBC instances are lacking up front documentation to let people know what to expect, and what to do with the blank canvas of their brand new WBC instance. If they experienced WD first, the contrast of having access to all WD provides, including tools, gadgets and a lot of documentation, versus what is available in a WBC instance makes that blank canvas a bit more daunting.

Opportunities:

Before a Cloud is created, provide easily accessible and broadly understandable information about what WBC instance is, what one needs to know to be successful, and what to expect after creating a brand new WBC instance. This will reduce the challenge of confronting a blank canvas with a new instance, and inform people what to expect after creating one. Providing a description of what features and functionality WBC instances have compared to WD and WBS would be helpful for admins to make decisions. Including basic, high level descriptions of the concepts that people need to understand and where to learn more will reduce the time it takes for people to know they need to know more, and where to find that important information. Describing the high level steps WBC admins need to do to prepare their WBC instances for housing the knowledge they and their community or institution hold will support admins to plan and prepare accordingly. Include connection to existing support resources like the Telegram group, and the Wikibase User or Stakeholder groups in order to relieve the need to discover and find where to get help.

To bolster support for admins once they have created a WBC instance, there is an opportunity to provide more detailed education about the concepts that need to be understood and resources for learning how to apply those concepts technically, in their instance. In context support reduces the time and effort to find help outside the workflow one needs help for. If it is understood where WBC admins need help as they prepare their instance for housing knowledge, the help can be added where it is needed. Along with providing an instance of a WBC, provide clear, in context education in support of admins understanding the concepts required along with the technical tasks they will need to accomplish preparing the WBC instance for ingesting and sharing the knowledge they hold.

Suggestions:

- Learn, from both admins who have created and prepared WBC instances, and from people who are in the process of creating and preparing WBC instances, what kind of support they needed along the process. Discover at which points in the process of preparing their instance for data ingestion and use, did they need help and support, and what the topics were. Identify what kind of conceptual or technical learning is needed at which points in the process, and design ways to provide it in the context of a WBC instance. Reduce the need for WBC admins to seek and find what they need to know to proceed with their instance.

WBC Finding 3. **Serving as a bridge between technical and knowledge expertise, admins need more support**

The need for a person to bridge technical expertise and knowledge expertise to support communities and institutions using WBC, combined with the lack of an intuitive user experience, creates more work for WBC admins. Though setting up a Cloud instance can be accomplished easily in a matter of minutes by anyone who can click a button (even people who are not experienced with using Wikibases), WBC instances are not intuitive to structure properly, to fill with data, and to be used by non technical people. WBC instances are not very customizable, and only come with the query service, the bulk import service, and a few extensions. WBC also lacks some tools that are standard in WD that make it useful for a community or institution to use, and, importantly, useful for reducing the work of admins.

The knowledge experts, or people who hold the knowledge to be shared, may not be experts, or confident in using the technology of a Cloud instance. Having no experience with creating software or making technology work, those people may find it scary and even intimidating to try, much less learn to use the technology without support.

The lack of automation and intuitively designed user interfaces and user experience in WBC instances makes it hard for the people who create Cloud instances to invite their communities or team members to participate. Before their instance can be populated with data, admins need to conduct training about the important concepts underlying the LOD ecosystem, and how to use WBC to both share knowledge and find and use knowledge. Admins might also create queries ahead of time to enable their communities to query and make use of the data without needing to know how to use the query service.

"I couldn't begin building out the community of editors yet. Because right now, there needs to be more in the Wikibase. You need to hit a certain point with enough content in a Wikibase to convince other people to begin coming in. You have to have something you can present to them so they are impressed enough that they're willing to take their time and their energy to also contribute " - Participant

In some cases, admins do data entry themselves after designing an accessible, friendly and structured way for people in their community or institution to share their knowledge with the admin rather than directly into their instance. It may currently be easier for admins to do the work of data ingestion into their Cloud instance than to set up for, and train people to directly share the knowledge they hold themselves.

In some cases, admins design an easy and structured way for people in their community or institution to share their knowledge with the admin rather than directly into their Cloud instance, perhaps in a spreadsheet. Then the admin may do the data entry themselves. It may currently be easier for admins to do the work of data ingestion into their cloud than to set up for, and train people to directly share the knowledge they hold into their instance.

A number of participants are doing the work of setting up and training for non technical people, and people not familiar with the Wikimedia movement and its technologies. One participant described their work, across several projects, to support their community in sharing the knowledge that community holds. The projects were focused on collecting and contributing photos of buildings in their country, names of places in their country in their own language, and recordings of the proper pronunciation of words in their language. During the first project, this participant learned that non technical people in their community were having difficulty contributing their knowledge directly into their WBC instance. So, the participant designed a way for the people in their community to contribute the relevant data to a spreadsheet. This was successful in gathering the community's local knowledge, as people knew how to use a spreadsheet. Then, the participant and their team ensured proper organization of the data in the spreadsheet and then uploaded and contributed the data into both their WBC instance and WD. The participant noted the usefulness of a Google add on for QuickStatements for this work.

There were a few other examples of WBC admins training the people who hold, or know the knowledge, to contribute directly to a WBC instance. For example, training librarians to contribute the publications in their library, or training historians working on a big project together how to accurately contribute data about the histories they are researching and sharing into a WBC instance. Training was described by one participant as needing to go slowly and one step at a time for the people who are new to Wikibases.

Opportunities:

Better support WBC admins by providing up front educational materials about the concepts required to use a WBC instance. This will support admins in understanding the concepts required and serve to support their work in educating and training the people they administer their instance for who are experts in, and will contribute the knowledge held.

Invite people without technical backgrounds, who are experts in the knowledge they hold, to easily contribute their knowledge to a WBC instance by designing and providing more intuitive contribution workflows. Different skill sets are required for admins to prepare a WBC instance than are required for people to contribute knowledge to that cloud after it has been set up. Considering these differences, and providing support for both kinds of users' needs could ease admins workload and inspire and support knowledge experts to contribute to WBC instances more easily.

Suggestions:

- Conduct research to learn from WBC admins, what they need to do to make their instances intuitive and accessible for their communities or teammates to use for both contributing knowledge and querying and using knowledge. Use this learning to design and provide more intuitive user experiences considering there will potentially be many people without technical backgrounds wanting to use WBC instances to share their knowledge.
- Learn about the various non technical ways WBC admins have invited people to share their knowledge into any of the Wikimedia LOD products and share these tactics with anyone creating a WBC instance.

WBC Finding 4. Wikibase Cloud's lack of functionality makes a challenging start

WBC's lack of functionality compared to WD makes for a challenging start especially for those who were introduced to the LOD ecosystem through WD. Participants reported a lack of features and

functionality in WBC instances that they need for their work. These items were consistently identified as needs when preparing to populate their instances with the particular knowledge they are trying to share:

- No multimedia hosting to support visual knowledge (and for supporting sign language)
- Lack of javascript to create intuitive user interfaces and a user experience
- Difficulty in adding languages with no ISO code

Notably, the difficulty in adding languages to a WBC instance is an area to seriously consider. Marginalization can happen through a lack of language support on the internet. A powerful way to reduce marginalization is to provide easy access to functionality for as many of the 7000 languages that humans use to communicate as possible.

Opportunity:

Provide the same capabilities available in WD in WBC instances to make admins' work easier, and reduce the time they need to make their WBC instance work as required. Including the same capabilities, gadgets and tools available in WD could make it easier for admins to prepare their WBC instances for ingesting the knowledge their communities or institutions hold. In some cases, admins use WBS to add capabilities, like multimedia hosting for example, to their WBC instances. Not everyone would have the technical skills to accomplish this. Providing more capabilities, tools and gadgets from the beginning could make it easier for a broader range of people to more easily prepare their instance for ingesting the knowledge their community or institution holds.

Suggestions:

- Learn from WBC admins what features they need, or had to create in their WBC instances to prepare it for use, and prioritize provision as possible.
- Create and make available the documentation and support needed for adding the requested features in an instance if it is not possible to provide those functions upon set up.

Requests:

- Two participants suggested automating the workflow required to add a language to a WBC instance. It was also suggested to automate the workflow for creating and adding new lexemes. It is possible to add languages now, the current workflow is complex and time consuming. People need to research, find support, and learn a lot to accomplish adding a language, or for creating new lexemes.
- Add the local WBC instance's prefix to the list of standard prefixes in a WBC query service. This helps people query without having to dig in and learn why their query isn't querying their instance.
- A request was made to have a version of Cradle that can be used on WBC instances - not referring to WD.
- Several participants directly requested data validation for WBC instances. They want a tool that validates data as it goes in so they don't have to fix it later.
- Two participants suggested a bot to look at all links used as references and check if the links are still live or not. And, a language model along with the bot that assesses the references asking if the claim made in argument is supported by the reference. Verification of references and an easy way to repair broken links to references is the overarching request.
- Below is a list of the features and functionality that were requested by participants:

- Multimedia hosting particularly for using lingua libre in a WBC and the ability to use sign language (images)
- The same more performant Query Service updaters that WD has (the old one on WBC is underperforming)
- Cradle that doesn't automatically refer to WD
- Data validation (error checking) on ingestion (saves a lot of time and work for people, and reduces risk of causing others' work in fixing errors)
- Depicts for creating structured data for media
- Infoboxes: an easy way to create and use info boxes that are not connected to WD, but used in a WBC instance or a self hosted Wikibase.
- Trigger warnings: a feature that can be used in a Wikibase to warn people (who are taking in content of that Wikibase) that there is content that may be triggering ahead. This would serve to provide an opportunity for folks who don't want to experience potentially triggering content, to avoid it. **Note:** This is possible for an admin to do on a particular Wiki by adding a site notice, but not everyone may know it exists.
- Automation of data dumps for backups
- Automation to find and repair broken links

WBC Finding 5. **Wikibase Cloud admins need support for federation**

WBC admins need education and support to make federation of Wikibases possible. Though Wikibase federation is emergent and is hard to accomplish at this point, a high level vision of federation as a way to combine knowledge from more than one Wikibase instance into a big picture of knowledge from many perspectives was described by some participants. Federation brings up the idea of a plurality of knowledges existing together in the Wikimedia LOD ecosystem, each holding authoritative knowledge and sharing it back to WD and other Wikibases.

“Considering a strategy to address the problems with P21, I could make a Wikibase Cloud, create a policy, work with others on it, showcase it on that Wikibase Cloud and hope it will be adopted in Wikidata at some point.”
-Participant

“The vision is of minoritized activists seeing us [Wikimedians] as part of their strategies for breaking silences and challenging power - not only us trying to sell them a solution.” - Participant

WBC instances are considered by some as a home for marginalized knowledge, and a safe place for people who hold that knowledge to safely contribute knowledge, and not face the kind of challenges faced in WD when attempting to share some kinds of knowledge.

“Additionally, we have a vision that marginalized people feel that sharing knowledge is something that they can do safely and happily - not only as passive subjects or objects of knowledge of others.” - Participant

“We want to create space in the open knowledge framework and the open data framework not just about operations, but focusing on joy and humanity too.” Participant

To accomplish the vision of federation, people participating in the LOD ecosystem need to know about this vision and the value it could provide. People, especially admins, need to know how to work toward accomplishing federation in the best way for their community and knowledge. Federation takes work and can be hard to learn how to do it well. Admins need to learn how to map data between Wikibases (including WD) without causing problems for either Wikibase. Designing a SPARQL query, and referring it to the right bases, is challenging. One participant wanted to not have a query refer to WD, but was not able to figure that out, for example.

Opportunity:

Communicate clearly about the idea of and potential value of federation of Wikibases. Greater awareness of the idea of federation, and the potential for a plurality of knowledges to exist together, could serve as a motivation for people to use a WBC instance to share their knowledge, especially if it is marginalized knowledge. Just being aware of this possibility, even if federation is in nascent stages of being broadly possible and available, may bring more people who hold marginalized knowledge to participate in the Wikimedia LOD ecosystem.

Suggestions:

- Conduct a survey or other research to gauge WBC and WBS users' awareness of and understanding of the potential value of Wikibases being interoperable.
- Learn from WBC admins who have made a WBC instance interoperable with other Wikibases, what it took for them to learn how and what parts of the process were particularly difficult or hard to learn about. Provide support for admins to better understand and create the conditions for federation.

Wikibase Suite (WBS) and knowledge equity

Because of a lack of functionality, sometimes WBC doesn't suffice for the project at hand. In this case people may consider self-hosting a Wikibase using WBS. This choice is dependent on the person and their project having the money, the skills required to self express knowledge, and time to install a Wikibase and set it up to work for their project. The learning curve is currently steep, even for experienced software admins.

WBS provides an opportunity to experiment and test the powerful tools. However, participants describe needing different and more resources to use WBS, like time to learn, and money for hosting or hardware, than are needed for using WD or WBC. Though participants described needing time to learn to use WD, and to prepare a WBC instance to ingest knowledge, it takes more time to learn how to install and prepare a Wikibase using WBS. Like in WBC instances, it is necessary to design and build a user interface easy enough for non technical people to use once the Wikibase is installed on a server. Participants described needing to learn technical skills to accomplish both hosting and creating a usable Wikibase with what is provided in WBS. Participants also note that compared to the tools, gadgets and

capabilities of WD, WBS is missing gadgets and functionality that save time and effort. Providing more self-serve documentation describing conceptual and technical skills required, and including gadgets and tools like exist in WD, would go a long way in better supporting, and reducing the time it takes WBS admins to accomplish their work.

WBS Finding 1. **Resource gap to self-host a Wikibase**

Time and money is needed to self-host a Wikibase using WBS, and this creates a potential barrier for those wanting to share marginalized knowledge. Setting up a Wikibase on one's own server requires time and budget for hosting. Either there needs to be a person with the technical capability, the time, and budget to self host a Wikibase, or there needs to be someone with the time and patience to learn how to do the technical work of self hosting a Wikibase, as well as the budget for hosting. There are services that will do the set up for self hosting a Wikibase for a fee.

Though it is possible for people with no background in technology to self-host a Wikibase using WBS, it can take a long time to learn how to do so and to prepare their Wikibase for use.

“I have a background as an artist and a designer. It's my education. I was able to understand linked open data for the first time. And I've seen Wikibase before. It was just some gobbledygook. I think that's an incredible quality [that people with non technology backgrounds can learn to use Wikimedia software]. I think it could be also a chance for Wikibase to leap ahead and make this really easy to understand and to use for data curators.” - Participant

In one case, it took so long to find funding for, and to set up a Wikibase using Suite, that a new team of people needed to be educated and brought up to speed with using WBS in their organization by the time it was close to ready to use.

It takes resources, like time, hardware, budget to rent a hosting service, or to hire someone to set up self hosting using WBS. Smaller or less resourced organizations may not have the resources to use WBS or to hire someone to set up a Wikibase on their server as well as prepare their Wikibase for people in the organization or community to easily use. People may have the ability to learn how to use WBS, but not the time to do so. This is important because, as described by a few participants, organizations holding marginalized knowledge may have more of a tendency to be under-resourced.

“Having a system admin that can work in a collective -in Uruguay or Uganda, for example, is so hard. Such professionals already have a job, and they have little time to contribute for social movements.”
- Participant

“It is hard for groups situated in majority world territories to practice data sovereignty if they don't have access to manage infrastructure. We know people who have knowledge sovereignty and don't have the ability, and can't take on the risk of managing tech infrastructures.” - Participant

Opportunity:

Identify or create support services and structures particularly for people who want to self host a Wikibase using WBS for under-resourced communities and organizations in order to make using WBS more broadly available for housing and sharing a variety of knowledges. Because funding can be a blocker, grants or other creative ways of providing technical and financial support needed to use WBS could broaden access to under-resourced communities and organizations who want to contribute their knowledge to the Wikimedia LOD ecosystem. Since the time it takes to learn the technical capabilities for installing a Wikibase and self hosting can be extended, perhaps providing a service for installing and setting up self hosting a Wikibase could be identified or provided for organizations or communities that are under resourced. This may also broaden access to a wider diversity of communities and organizations.

Suggestions:

- Aware of Wikimedia's funding programs that provide funds for Wikimedia projects, one participant suggests the idea of having a similar program where the "award" would be developer time and technical support in lieu of funds. The participant notes drawing inspiration from [Rhizomes's 7x7 program](#) which pairs artists with technologists to build something together.
- One participant suggested having staff serve as “Wikibase as a Service”, a unit to install the software for a nominal service fee.

Requests:

- Work with under-resourced organizations to find funding or extra support to get their Wikibases up and running.
- Wikibases are designed and set up for one person to interact with the system at a time. Many end points all doing individual contributions and work. Since many communities work in collective, collaborative ways, how could Wikimedia products support communities and groups who make decisions and work as a group?

WBS Finding 2. Learning curve to understand complex concepts and specialized technical skills

Specialized technical skills and understanding complex concepts is required to use WBS for a Wikibase, potentially creating a long time horizon for set up. Even if someone has a background in technology or has become familiar with WD and has set up a WBC instance for a project, WBS is more complex to use than either of the other products. This is partially due to Mediawiki software being different from other software people might be experienced installing, and WBS having little documentation to directly guide admins through the process of set up and implementation of their Wikibase instance.

Like with WBC, there sometimes needs to be a person to act as a bridge between the people who are experts in the knowledge being shared and the technology being used. Sometimes this person doesn't have a technical background. As one participant described:

"Subject experts may need limited training to become data specialists who curate descriptive statements. But possessing technological know-how of WB is a whole nother story." - Participant

As is the case with WBC, an understanding of concepts like linked open data, data modeling and what an ontology is, is needed to properly describe the knowledge in a Wikibase, once it is installed. Some things, like concepts of time different from the Gregorian or Julian calendar, for example, can be very hard to model in a Wikibase. People working in institutions, museums and libraries may be beholden to certain types of data models by their organizations, such as authority files created specifically for certain collections or knowledges. These kinds of challenges are both conceptual challenges about understanding how to model data that isn't easily modeled using Mediawiki software, and a technical challenge in understanding how, technically, to map a model from a database to a Wikibase. There are people skilled at data modeling and mapping who try and are able to map complex concepts like these, but it is not easy. Not everyone would have the high level understanding of data modeling and the technical chops and tenacity to map the model back to a Wikibase in order to accomplish such a feat.

Opportunity:

Consider providing documentation and education usable by both people who have technical backgrounds and people who don't have a technical background to enable less lengthy and challenging installation processes using WBS. Experienced technologists who perhaps have learned their skills using commercial software may need different support than people with no technological background or experience self hosting software. Along the path of installing a Wikibase on a server, provide detailed information on how to proceed and what options are available for self hosting a Wikibase.

Suggestions:

- Learn from WBS users who are currently self hosting, and those who are in the process of self hosting a Wikibase, what challenges they encountered along the path they took in self hosting a Wikibase using WBS. Map the patterns of challenges described by participants to the process of self hosting and preparing a Wikibase for use. Provide important information that is currently missing, along the path of production, so WBS users don't need to go and find support for known challenges that already have answers.

WBS Finding 3. Skills required to self express knowledge

The need to create a user experience for people who want to share the knowledge they hold, takes time and requires design and engineering skills of admins who prepare a Wikibase for use by their communities, organizations and institutions. The greater ability to customize a Wikibase using Suite is a big draw to the service. This is because it allows admins to add important features not available in a WBC instance like multimedia hosting and useful gadgets and extensions that make Wikibase workflows more efficient and easy to conduct. Importantly, unlike WBC instances, Javascript

can be used with WBS for customizing the user interface and the user experience. Some participants reported using WBS to add features and functionality to their Cloud instance. Most prominently, participants used WBS to add multimedia hosting to their WBC instance.

However, the ability to customize takes time and design and engineering skills. After someone successfully uses Suite to install a Wikibase on a server, they are then confronted with a blank page. There is no guidance on next steps on how to add the features they are excited to implement, nor is there education about the possibilities of what can be done with that powerful software. As described by participants, it can be hard to know what to do next, and how to set up a Wikibase to accomplish a vision with little to no documentation.

For example, without education about ontology and documentation describing how to model particular data types, a person may set up their Wikibase with a standard Western ontology like WD uses instead of creating an ontology that accurately reflects the kind of data they need to model. Some participants described using trial and error and learning as they went to learn what they needed to for self hosting. This kind of learning takes more time from potentially under-resourced organizations, and from people who may be performing two roles in their job. One role is their chosen profession - perhaps librarianship, a form of research, or advocacy for a community - and the second role is being a technologist to support their community or team by creating, setting up and maintaining their Wikibase. Some participants report a fuzzy line between volunteer work and the work of their job when advocating for and conducting their administration of a Wikibase. Some participants are fully volunteering their work on Wikibases and providing value to their communities and organizations without compensation. When this is the case, people still need to make a living. In this case, work on a Wikibase may be deprioritized because making a living is obviously a priority.

The people who are experts in the knowledge being shared in the Wikibase, if they aren't librarians, may not be fluent in data modeling, creating ontologies or schema, nor may they be experienced in writing SPARQL queries. Because of this and a lack of a standard user interface and user experience in a Wikibase, admins need to do the work not only to set up and start hosting a Wikibase, but they need to to prepare their Wikibase to make it work for the experts in the knowledge being shared. Admins would benefit from WBS providing an intuitive user interface and user experience to support the people who are knowledge experts in participating in knowledge sharing using their Wikibase because it would save them time and work.

Opportunity:

Design and provide as much pre-loaded functionality, including intuitive user interfaces and experiences, within WBS as possible to reduce the time and learning burden on admins in customizing and preparing their Wikibase for use. To start, provide the features that come along with WBC, and also consider providing what useful tools and gadgets that exist in WD, within WBC. Support WBS admins with the documentation and know how to appropriately customize their Wikibase for the knowledge it will house. Include examples of different ontologies and education about variance of ontologies and how to create them in a Wikibase.

Suggestions:

- A few participants described the idea of providing a full featured modeling tool providing pre-loaded ontologies (especially of existing models that could be reused), and schemas as a form to start people out in creating their Wikibases because the blank page issue makes it hard to get started.
- An a la carte service for WBS so admins can see a kind of menu of Wikibases to choose from. For example, a Wikibase with name authority control for people, or with particular ontologies used in specific kinds of data sets. Then admins could more quickly and easily have a baseline set up for their Wikibase using WBS.
- One participant suggested speech recognition be used to accomplish tasks in Wikibases to address the lack of intuitive user interfaces and user experiences.

Requests:

- There are more features provided ready to use in a WBC instance. Add those same capabilities to WBS upon install at a minimum.

WBS Finding 4. System functions standard in WD, missing in WBS

System functions standard in WD are missing in WBS, creating more work for WBS users as they prepare their Wikibase. WBS is a powerful bunch of technology. It can be used to customize a Wikibase to appropriately model, ingest and query knowledge. It can be challenging to set up a Wikibase with Suite because first, important tools and gadgets need to be found and set up. There is little to no documentation describing where to find and how to install them, and even little education on the existence of these tools. Participants report noticing the lack of installed gadgets and functionality in WBS that are used in WD to both automate and make workflows more efficient. The lack of pre-installed gadgets and tools increases admins' workload, and is another layer of work for admins to find, install, and make sure these tools and gadgets are working in their Wikibase. For example: there is no link for the query service, Cradle needs to be installed, and other basic gadgets that people use in WD also need to be found and installed. Another reported issue is difficulty in accessing Wikibase updates through Docker. It takes a lot of time and effort, and because of this, updating a Wikibase through Docker can be delayed by busy admins, perhaps not being paid for their efforts.

Opportunity:

Better support WBS admins by preloading gadgets, tools and capabilities that are known to save time and effort and increase the quality of data in WD and in WBC. Setting up WBS admins with these capabilities without needing to find and install them will reduce the amount of time needed and the learning curves participants described. In turn, WBS admins could use their time and focus their attention on customizations that are needed only in their Wikibase, like creating an ontology specifically for the knowledge they will house and share in their Wikibase. This could reduce the time it takes for WBS users to succeed in self hosting and engaging their community or organization to contribute and make use of their Wikibase. It may also better support under-resourced organizations and communities to use WBS.

Suggestions:

- Poll or talk with WBS admins to learn about the features and functionality they need to be successful at self hosting and for setting up their Wikibases for ingesting and sharing

knowledge. Use what is learned to prioritize provision of the most important features and functionality according to people who have, or are self hosting a Wikibase.

WBS Finding 5. **Federation of Wikibases is a compelling vision**

Federation of Wikibases housing and sharing a plurality of human knowledges is a compelling vision, especially if people are aware of the idea and understand it. As described by participants, it can be challenging to understand the concept of federation. Participants ranged from not knowing about federation at all, to having a vision of what federation could be without knowing how to accomplish it, to having accomplished federation of their Wikibase to a degree (after a lot of learning and trying), and everything in between. As described by participants, it can be difficult to understand how to properly federate a Wikibase so the Wikibase is representative of the data it holds and, at the same time, the Wikibase is interoperable with other Wikibases. Clear and easy to find education and documentation, missing now, could support accomplishing the vision of federation.

Because of the ability to choose what and where to query in the Wikimedia LOD ecosystem, a user of a Wikibase can choose what data to query and ingest into their Wikibase. There is also the ability to choose what not to query for and what data to not take into a Wikibase from another Wikibase. For example, from their self hosted Wikibase, a person could query WD for biographies of a specific group of people and choose not to include Property 21 (sex and gender) in the query. In this way, data can be shared the way it needs to be for the particular quarant's needs. Bringing more people along with the idea and vision of federation could add to the ability to innovate and proceed with making the vision a reality.

Opportunity:

Collaborate with Wikibase users on what federation of Wikibases could provide for the expression of a plurality of knowledge, and how that might benefit humanity. Understand what needs to happen in more detail, for the vision to be accomplished technically, and to be made accessible for a broad population, of variously experienced people. There is an opportunity to share the vision and value of Wikibase federation more broadly when people will be able to make use of it more easily.

Suggestions:

- Learn from people who have worked to understand how to federate, or make Wikibases interoperable, perhaps people who participated in beta testing for federation, what their visions and imagined possibilities for federation are. Because it is an emerging capability, learning from those who are trying could be useful to learn how to communicate about the possibility and how to accomplish that vision.

Conclusion

Together, the products of the Wikimedia LOD ecosystem; WD, WBC and WBS, have the potential to share a greater plurality of human knowledges accessible to anyone on the internet. Since 2012 when WD first arrived, to provide info boxes (details about a subject of an article) for Wikipedia and other Wikimedia projects, it has evolved to a single linked open data repository, housing multitudes of knowledge, including marginalized knowledges, and to a nascent ecosystem of Wikibase instances sharing more knowledge than could be afforded solely in Wikipedia and WD. Designed for humans and machines to work together, the software provided by Wikimedia Deutschland offers an opportunity for everyone who can access the internet to share the particular knowledge they hold, and perhaps even in their language.

Like people, communities, and knowledge in the physical world, knowledge and histories have been erased, made invisible and marginalized on the internet. For example, there are 7000 human languages and only about [300 of our languages are available](#) to use easily on the internet. This is a form of marginalization. For someone to make use of and contribute to the value of the internet, it is required to know one of the dominant languages that is supported on the internet. This requires some people, whose first language is not fully supported on the internet, to do more work and take more time to participate than a person whose first language is one of the fully supported languages. Consider the communities whose knowledge is shared orally with language that is not written. How does one share their knowledge to the Wikimedia LOD ecosystem in this case?

This project is a part of, and builds on years of research, structured data decolonization efforts, and community work within and around the Wikimedia movement to welcome more kinds of knowledge than can be afforded in Wikipedia and Wikidata. This project's contribution starts with the understanding and acknowledgement that knowledge equity is not yet accomplished in the LOD ecosystem. It continues in support of the intact vision of communities and data in the LOD ecosystem representing the world in an equitable way. Toward accomplishing that vision, this report provides opportunities for the Wikimedia Deutschland product teams to address and reduce barriers to knowledge equity and inspiration for building infrastructures to better foster knowledge equity.

WD provides enormous value for humanity as the world's largest open data repository. It can make the invisible, visible, and make it easy for people and machines to find and make meaning of knowledge that was previously difficult to find. This may be a particularly compelling opportunity for people who hold, and want to contribute knowledge that has been marginalized or erased. The people who participated in this study, most of whom contribute historically marginalized knowledge, shared stories about their optimism about the Wikimedia movement, and their use of WD, WBC and WBS to structure and share their knowledge accurately. They also shared the challenges they experience in the LOD ecosystem, as they attempt to accurately describe the knowledge they hold.

Each product contributes differently to people's ability to share the knowledge they hold, whether marginalized or not. These products provide capabilities to make knowledge widely accessible and findable, and to co-exist with the rest of humanity's knowledges. Though Wikibase federation is emergent and is currently hard to accomplish, a high level vision of federation as a way to combine knowledge from more than one Wikibase instance into a big picture of knowledge from many perspectives was described by some participants. Federation brings up the idea of a plurality of

knowledges existing together in the Wikimedia Linked Open Data ecosystem, each holding authoritative knowledge and sharing it back to Wikidata and other Wikibases.

This project would not have been possible without the participants' generosity with their time, openness in sharing their perspectives and experiences with the products, and their patience in explaining technical details and concepts about the LOD ecosystem. The collaboration of Annie Kim and Jon Amar in defining, accomplishing and nurturing the project, along with the Linked Open Data teams, Communications and Partnerships teams support and collaboration made this work possible. Thank you. Let's keep moving towards knowledge equity in the Linked Open Data ecosystem!

Appendix

For Wikidata Finding 1

Additional information a participant's idea to more **easily find the correct properties in the Wikidata ontology**: Here is a screenshot of how the participant thematically organized the properties it is easier to find the appropriate property for an item:

- **Country** : head of government (P6), continent (P30), head of state (P35), capital (P36), official language (P37), currency (P38), flag image (P41), shares border with (P47), top-level Internet domain (P78), anthem (P85), basic form of government (P122), flag (P153), country calling code (P474), diplomatic relation (P530), central bank/issuer (P562)
 - External ID : ISO 3166-1 alpha-2 code (P297) (2 letter), ISO 3166-1 alpha-3 code (P298) (3 letter), ISO 3166-1 numeric code (P299), ISO 3166-2 code (P300) (2 letter + subdivisions), ISO 4217 code (P498) (currency code)
- **General datatype** : video (P10), image (P18), audio (P51)
- **Image-related properties** : traffic sign (P14), coat of arms image (P94), signature (P109), chemical structure (P117), logo image (P154), seal image (P158)
- **Person** : place of birth (P19), place of death (P20), sex or gender (P21), country of citizenship (P27), position held (P39), family (P53), ancestral home (P66), educated at (P69), sexual orientation (P91), noble title (P97), field of work (P101), member of political party (P102), native language (P103), occupation (P106), signature (P109), place of burial (P119), religion or worldview (P140), ethnic group (P172), official residence (P263), military rank (P410), canonization status (P411), voice type (P412), position played on team / speciality (P413), shooting handedness (P423), botanist author abbreviation (P428), astronaut mission (P450), character role (P453), dan/kyu rank (P468), cause of death (P509), honorific prefix (P511), academic degree (P512), chivalric order (P550), residence (P551), handedness (P552), date of birth (P569), date of death (P570), conflict (P607)
 - Relationship with other person : father (P22), mother (P25), spouse (P26), child (P40), killed by (P157), doctoral advisor (P184), doctoral student (P185), unmarried partner (P451)
 - Organization (as an object) : member of sports team (P54), employer (P108), military branch (P241)
 - External ID : Find a Grave memorial ID (P535), ATP player ID (P536), Mathematics Genealogy Project ID (P548), IPNI author ID (P586) (international plant name index), WTA player ID (P597) (female tennis player), Léonore ID (P640), RKDartists ID (P650), Biografisch Portaal van Nederland ID (P651)
- **Organization (as a subject)** : main regulatory text (P92), founded by (P112), home venue (P115), logo image (P154), seal image (P158), headquarters location (P159), chief executive officer (P169), legislative body (P194), business division (P199), executive body (P208), highest judicial authority (P209), coat of arms (P237), ticker symbol (P249), has subsidiary (P355), stock exchange (P414), has seal, badge, or sigil (P418), industry (P452), foundational text (P457), chairperson (P488), inception (P571), dissolved, abolished or demolished date (P576), religious order (P611), mother house (P612)
- **Upper-level ontology** : instance of (P31), location (P276), subclass of (P279), said to be the same as (P460), member of (P463)
- **Works** : author (P50), director (P57), screenwriter (P58), architect (P84), composer (P86), librettist (P87), commissioned by (P88), editor (P98), illustrator (P110), publisher (P123), genre (P136), architectural style (P149), cast member (P161), producer (P162), creator (P170), performer (P175), manufacturer (P176), developer (P178), part of the series (P179), depicts (P180), made from material (P186), main building contractor (P193), record label (P264), production company (P272), copyright license (P275), place of publication (P291), director of photography (P344), original language of film or TV show (P364), presenter (P371), soundtrack release (P406), language of work or name (P407), distribution format (P437), review score (P444), review score by (P447), original broadcaster (P449), volume (P478), recorded at studio or venue (P483), exhibition history (P608), edition or translation of (P629), translator (P655), tracklist (P658), characters (P674), lyrics by (P676)
 - External ID : Union List of Artist Names ID (P245) (Getty Research Institute, visual arts), IMDb ID (P345), Joconde work ID (P347) (French), RKDimages ID (P350) (Netherlands), MusicBrainz artist ID (P434), MusicBrainz work ID (P435), MusicBrainz release group ID (P436), FilmAffinity film ID (P480), Palissy ID (P481) (moveable objects of French cultural heritage), cultural properties of Belarus reference number (P632), Quebec cultural heritage directory ID (P633)
- **Sky objects** : constellation (P59), site of astronomical discovery (P65), minor planet group (P196), spectral class (P215), galaxy morphological type (P223), astronomic symbol image (P367), space launch vehicle (P373), located on astronomical body (P376), parent astronomical body (P397), child astronomical body (P398), companion of (P399), provisional designation (P496), orbit diagram (P491), type of orbit (P522), time of object orbit decay (P621)
 - External ID : COSPAR ID (P247) (satellite, UN Committee on Space Research), SCN (P377) (Satellite Catalog Number)
- **Things** : discoverer or inventor (P61), maintained by (P126), owned by (P127), operator (P137), named after (P138), award received (P166), location of discovery (P189), designed by (P287), country of origin (P495), time of discovery or invention (P575)
- **Transportation system** : transport network (P16), connecting line (P81), airline hub (P113), airline alliance (P114), adjacent station (P197), station code (P296), callsign of airline (P432), scheduled service destination (P521), runway (P529), terminus (P559), direction (P560), terminus location (P609)
 - External ID : IATA airline designator (P239) (airline), ICAO airline designator (P230) (airline), IATA airport code (P238), ICAO airport code (P239), FAA airport code (P240) (US only)
- **Species** : taxon rank (P105), IUCN conservation status (P141), parent taxon (P171), taxon range map image (P181), endemic to (P183), taxon name (P225), taxon author (P405), taxonomic type (P427), basionym (P566), year of publication of scientific name for taxon (P574), *incertae sedis* (P678), replaced synonym (for nom. nov.) (P694), ex taxon author (P697)
 - External ID : EE breed number (P303), IUCN taxon ID (P627), NCBI taxonomy ID (P685) (taxon), BHL page ID (P687) (biodiversity heritage library)
- **Chemicals** : chemical structure (P117), regulates (molecular biology) (P128), canonical SMILES (P233), InChI (P234), InChIKey (P235), element symbol (P246), chemical formula (P274), interleaves with (P514), phase of matter (P515), streak color (P534), twinning (P537), mineral fracture (P538), crystal system (P556), crystal habit (P565), point group (P589), route of administration (P636), genomic start (P644), genomic end (P645), genomic assembly (P659), EC enzyme classification (P660), molecular function (P680), biological process (P682), ortholog (P684), encodes (P688), space group (P690), Gene Atlas image (P692), cleavage (P693), encoded by (P702), found in taxon (P703)
 - External ID : CAS Registry Number (P231), EC number (P232), ATC code (P267) (anatomical therapeutic chemical classification system), Entrez Gene ID (P251), UniProt protein ID (P352), HGNC gene symbol (P353), HGNC ID (P254) (HUGO Gene Nomenclature Committee), IMA Number, broad sense (P434) (mineral), IMA status and/or rank (P579) (mineral), EC enzyme number (P591), ChEMBL ID (P592) (bioactive molecules with drug-like properties), HomoloGene ID (P593), Ensembl gene ID (P594), Guide to Pharmacology Ligand ID (P595), E number (P628) (food additives), RefSeq protein ID (P637), PDB structure ID (P638) (protein 3D structure), RefSeq RNA ID (P639), UNII (P632) (FDA ingredient identifier), RefSeq (P656) (protein and RNA), RTECS number (P657) (Registry of Toxic Effects of Chemical Substances), ChemSpider ID (P661), PubChem CID (P662) (chemical molecules and their activities in biological assays), KEGG ID (P665) (genes and genomes), Mouse Genome Informatics ID (P671) (genes), ZVG number (P679) (occupational safety), ChEBI ID (P683) (chemical entities of biological interest), Gene Ontology ID (P686), UN number (P695) (hazardous substances), Kemler code (P700) (hazardous substances), Ensembl transcript ID (P704) (gene transcript), Ensembl protein ID (P705), Strunz 8th edition (series ID, updated) (P711) (minerals), Nickel-Strunz 9th edition (updated 2009) (P712) (minerals), Nickel-Strunz 10th ed., review of (9th ed./ 2009 update) (P713) (minerals), Dana 8th edition (P714) (minerals), DrugBank ID (P715)
- **Sports** : league (P118), head coach (P286), general manager (P505), doubles record (P555), singles record (P564), team captain (P634), drafted by (P647)
- **Places or geostuctures** : located in the administrative territorial entity (P131), contains the administrative territorial entity (P150), structure replaced by (P167), crosses (P177), twinned administrative body (P190), inflows (P200), outflows (P201), basin country (P205), located in or next to body of water (P206), bathymetry image (P207), locator map image (P242), postal code (P281), licence plate code (P395), mouth of the watercourse (P403), patron saint (P417), located in time zone (P421), occupant (P466), lake on watercourse (P469), local dialing code (P473), exclave of (P500), enclave within (P501), commemorates (P547), underlies (P567), overlies (P568), dissolved, abolished or demolished date (P576), highest point (P610), OS grid reference (P613), coordinate location (P625), structural engineer (P631), located on street (P669), house number (P670), located in/on

Jon Kintree's suggestions

Jon Kintree, a retired librarian, from St. Louis Public Library, of Missouri, United States, Planet Earth, requested to share his suggestions on a few topics, and his idea for a world congress. He gave permission to be identified and to share this information here.

Suggestions for making querying easier:

- “For example, including a few sample SPARQL queries in the beginning of the prompt to a language model can help the model do a better job of generating additional SPARQL queries.
- Also, including some instructions such as: take this step by step, check your work after every step, and be able to explain what you're doing for each step, can also improve the quality of the output.
- A SPARQL plugin like this might help make Wikibases more accessible to nontechnical people, especially if it can be plugged into other language models in addition to ChatGPT.
 - <https://community.openlinksw.com/t/sparql-plugin-for-chatgpt/3978>

Suggestions for making Wikibase / Mediawiki more usable on mobile devices:

- The [limware.ai](https://www.limware.ai) effort is going beyond research and prototypes to production ready plug and play open source models that are optimized for CPUs, and small enough to run on 16 GB of RAM. Some of their models are trained to create SQL queries. Maybe a similar process can be used to create SPARQL queries.
- Qualcomm has technology to optimize the performance of software on the Snapdragon 8 gen 2 and other processors.
- Oneplus might help create a script that installs Wikibase on their smartphones.
- BTW, the Virtuoso Universal Server looks interesting as a lightweight SPARQL endpoint that might run well on a 16 GB Oneplus 11 smartphone.
- Here is a suggestion. WMDE could evaluate the just-released Phi-3 mini language model, which has 3.8 billion parameters, on its performance as an interface to a Wikibase.
 - <https://arstechnica.com/information-technology/2024/04/microsofts-phi-3-shows-the-surprising-power-of-small-locally-run-ai-language-models/>
- That performance could possibly be improved with in-context learning by including several examples of SPARQL queries in the prompt to the model.
- The performance could possibly be further improved by fine tuning or adapting the Phi-3 mini model with datasets containing SPARQL queries, including federated Wikibases. It would really be neat if Phi-3 mini could generate such a training dataset for itself.

Using LLMs to create ontologies more easily

- “It seems to me that if a language model can learn how to translate between human languages, and to write computer code, that it can learn to translate between ontologies. So, a person can speak in any language, and the model can extract entities and relationships along with their attributes, and can generate queries that do entity resolution, graph merging and updating, reasoning, and ranking of statements from federated graphs for retrieval in conversations

More open source

- MIT has an Institute for Collective Intelligence. The Allen Institute for AI is exemplary in releasing open source models and datasets. The Hong Kong Institute of Science and Technology sounds good, too.

About using LLMs for getting answers (querying)

- <https://blog.langchain.dev/enhancing-rag-based-applications-accuracy-by-constructing-and-leveraging-knowledge-graphs/>

The idea of a world congress accessible and able to be participated in through smart phones, and ideas for accomplishing it:

- <https://youtu.be/ecU0BWzmDa0>
- <https://youtu.be/xzICL4v8SOW>
- IMHO, the following is the essence of the message that we want, in the words of Buckminster Fuller, "to get people in the know as quickly as possible."
 - Language models can represent entities and relationships in graph structures of items connected with properties that can become a shared world model for effective human and digital actions.
 - There is a concept of a Wikibase Ecosystem that could become such a shared world model. Wikidata is an example of a Wikibase. Language models can also become user friendly interfaces for the shared world model.
 - Smartphones have become powerful enough that some of the newest phones can play an important role in building a global platform for a shared world model, and for collective human and digital intelligence.
 - The principles of the Universal Declaration of Human Rights and the Earth Charter are among the strongest and most enduring examples of human intelligence. Those principles are a great place from which to plan our actions.
- Again, maybe the public awareness will come when we actually have language models running on smartphones that are building the shared world model.
- It doesn't have to be perfect. It just has to be good enough. As Earth Mama sings, "Good enough is as good as a feast." Once we have good enough, we can make it better.