Decolonizing the Internet's Structured Data

SUMMARY REPORT

IN OCTOBER 2021, OVER 40 PARTICIPANTS FROM AROUND THE WORLD WERE BROUGHT TOGETHER FOR A CONVERSATION ABOUT DECOLONIZING THE INTERNET'S STRUCTURED DATA, LEADING UP TO WIKIDATACON. THE EVENT WAS JOINTLY ORGANIZED BY WHOSE KNOWLEDGE?, WIKI MOVIMENTO BRASIL, AND WIKIMEDIA DEUTSCHLAND. THIS IS THE REPORT BACK FROM OUR GATHERING AND THE REFLECTIONS THAT EMERGED FROM IT.

Content

Why did we organize this conversation?

Structured data is at the core of how the internet - as we currently know it - works. These are pieces of information organized in such a way that they can be easily read, understood, and processed by machines. Countless apps and platforms are built upon such structured data systems, including tools created by content providers like Google and collaborative initiatives like Wikidata. As a result, they permeate every corner of our daily lives online, from maps and navigation systems to movie reviews and rankings.

Through these systems, massive amounts of data get sorted out, organized, and classified in relation to other pieces of data. As a result, they are informed and structured by and around specific regulations, traditions, and epistemologies. In an attempt to categorize the world, they prescribe certain frames and worldviews — in other words, they are far from the often-assumed neutrality of data. By analyzing the power dynamics of structured data, we can examine whose views, whose agenda, whose ontologies (categories of classification), and whose decisions build and sustain these classifications and systems. We can then work together to build more just and equitable systems of structured data.

This transformation of structured data systems - and knowledge itself - requires an ongoing conversation and commitment that is not limited to technology companies or specific sectors. It needs to be a collective effort from individuals, organizations, and communities with different forms of knowledge and expertise, a process we call allyship or "solidarity in action". Most importantly though, this shared process needs to center the expertise, lived experiences, and leadership of the minoritized majority of the world.

As we've shared before, only a fraction of online public knowledge is produced on or by women, people of color, LGBTQI folks, Indigenous communities, and peoples from the Global South, in languages that are not English or other colonial languages. Yet we are the majority of the world's population, the majority of those online, and the ones most impacted by how structured data is used, or abused. We are the communities that are most impacted by historical and ongoing structures of power and privilege and interlocking systems of oppression, whether it is colonization, patriarchy, racism, homophobia, classism, casteism, ableism, and beyond. This is why we need to start decolonizing the internet as a practice, not simply a metaphor: we need to transform the different forms of power and privilege that shape our digital technologies.

Learn more about why knowledge justice matters for structured data from this presentation by our co-director Anasuya Sengupta.

What a powerful and heartfelt presentation on knowledge decolonization and its importance to structured data from @Anasuyashh @WhoseKnowledge at #WikidataCon2021. Thank u. We are not here to fill the gaps, but to restruct and rethink the ways of doing things.

Flavia Doria
@aflaviadoria on Twitter, 29 October 2021

Decolonizing the Internet's Structured Data was born out of the pressing, urgent need for such a collective conversation, centering those who are often marginalized. On October 13th, 2021, we invited over 40 knowledge activists, community organizers, tech-builders, and other "unusual allies" to join a safe, multilingual, collaborative space. This group of thoughtful, powerful thinkers and doers, was mostly female-identifying (71%), in/from the Global South (66%), and indigenous/black/ people of color(s) (82%) in origin. The conversation served as a pre-conference for WikidataCon 2021, and was jointly organized by Whose Knowledge?, Wikimedia Deutschland, and Wiki Movimento Brasil. As a result, we also prioritized the participation of communities from Latin America and the Caribbean.

What did we do, and how did we do it?

Decolonizing the Internet's Structured Data was organized in three different panels. In Panel 1, our guest speakers offered *Perspectives and Provocations* to frame the overall conversations. Then, participants split into seven groups for discussions named *Imaginations and Implementations*. Finally, listeners brought back key insights in a session of *Listening and Learning*, in which we shared reflections and ideas for the next steps.

We started the conversation by clearly laying out three core beliefs and commitments: love, respect, and solidarity. Based on this set of guiding principles, we wanted participants to be aware of their positionalities and privileges, and to be able to be their full, multiple selves during the whole session. Our goal was to promote interactions centered around listening and learning, and around respect (of each other and of our time together). By engaging in meaningful conversations, we wanted to lay the groundwork for key areas and future concrete actions.

We carried out the conversation virtually, and offered simultaneous interpretation in English, Portuguese and Spanish. We wanted everyone to be able to fully express themselves in a language that they felt comfortable with and to share their experiences and knowledges with each other regardless of language barriers.

It has really been an extremely enlightening conversation, and listening to the voices from different academic disciplines, the dialogue of knowledges – which always enriches so much – and not only that but also from so many countries, and with so many languages. A little Babel.

MONICA CECILIA CALDERON
CLOSING SESSION, 13 OCTOBER 2021

Additionally, we established clear privacy practices in order to guarantee people's safety and wellbeing during the entire call. For instance, participants could keep their cameras off if preferred and needed to seek clear consent to publish and attribute any identifiable quote from another participant, before sharing it on social media during and after the event.

Here is how the conversations unfolded.

What emerged

Perspectives and provocations

The conversation began with a panel named *Perspectives and Provocations* in which we framed the urgent issues around decolonizing structured data, and panelists explained what it meant to them. This first part counted on the insights of Lydia Pintscher, who is a product manager at Wikidata; Matariki Williams, a Māori writer, curator and historian, who is currently at the Ministry of Culture and Heritage in New Zealand; Paz Peña, an activist and feminist techie based in Chile; and Stacy Allison-Cassin,

an academic at the University of Toronto, librarian, Wikidata enthusiast, and a citizen of the Métis Nation from Ontario, Canada. This initial panel grounded us in what is at stake when we talk about structured data.

We began by situating the questions and frames we would cover during the event: what structured data is, and what it means to build systems around it that center multiple epistemologies. In order to advance in those discussions, we were rooted in power analysis: questioning how the control over resources — from the design of platforms to the access to content in one's languages — plays out at different levels online. Questions such as "whose story?," "who is the storyteller and curator?," and "whose systems of categorization?" permeated many interventions. As the panel progressed, we engaged with the importance of structured data and laid the ground for imagining other possibilities.

In this first part of the event, panelists challenged the alleged neutrality of structured data. We framed structured data as pieces of ideology, not as neutral

categories that classify the world in a certain way. Participants navigated the nuances that emerge when we question how data is classified and based on which epistemologies, while also centering the needs and the right of refusal and determination of our communities. How do we build alternatives based on emancipation and liberation? As beautifully put by our panelists, this conversation needs to go beyond a nice-to-have, and be recognized as an urgent matter that requires resources in order to be changed and improved.

Here's a sneak peek of this initial discussion. Answers have been edited and condensed for the sake of clarity and length.

What does structured data mean, and why is it important that we talk about it?

We have all of these different systems that interact, that structure our movement through a space in particular ways, and sometimes that way of thinking really funnels us [...]. Each of those times when we encounter one of those systems of structuring it, there is a little piece of ideology that conceptualizes the world in a particular way. It is an interaction between those pieces of information or data and the information system that surrounds them. [...] So, structured data for me begins at that very fundamental place of how we structure the world around us, how we interact with different information systems. Then we come along to a system like Wikidata, which I think is really wonderful in a lot of ways, like its ability to be a collaboration between communities, and to have some agency in changing some of those knowledge structures.

Image by Whose Knowledge?, CC BY-SA 4.0, via Wikimedia Commons

— Stacy Allison-Casin

What does it mean to have multiple knowledge frames, multiple epistemic frames, or epistemologies at the heart of structured data?

One of the things I've just been thinking about is the way in which power plays a role in structured data. And as someone who is a user rather than a designer of structured data, especially as an indigenous person in a colonized country, the way that plays out in the areas in which I've worked is quite huge. [...] We do not fit within the system, as the system has not been designed by us. That has many implications in the way in which knowledge itself has been shared, or collected, and gathered, and added to over time. [...] When we talk about internal processes, from a Ministry perspective, it is about making and structuring data in a way that makes our internal work more efficient. But what does that mean for our partners? Are we then expecting them to fit within a template that enables our work, and are we no longer centering their aspirations on how they handle their own knowledge? [...] So, when it comes to thinking about approaching structured data from and with multiple epistemologies, it happens already, and it happens naturally as an indigenous person.

Image by Whose Knowledge?, CC BY-SA 4.0, via Wikimedia Commons

Matariki Williams —

How can we re-imagine structured data, especially from a feminist and anti-colonial lens?

The structuring of the data in these systems is a historical continuation of this white, Eurocentric, colonial, patriarchal fantasy, which assumes that the data floats. They float free from their origins, they are stripped from our bodies, from our history, and therefore they are free to navigate the world like a neutral currency, a universal currency. [...] I also believe that structured data often act as a hegemonic force, because they literally drag ways of seeing the world, as hegemonic forces do. This hegemonic force presents itself as inevitable, it seeks to be the only way to order the world. [...] If we want to see the data structured from a feminist perspective, we have to fight for the democratization of the discussion.

Image by Whose Knowledge?, CC BY-SA 4.0, via Wikimedia Commons

— Paz Peña

What is one thing you would like to see done differently in structured data today, that will help us come to that space of emancipation and liberation?

In Wikidata, we are trying to provide structured data in a way that is more just, more equitable, more representative of the different experiences and the world that we live in. [...] We're trying to strike a balance between providing useful and usable data that people can build applications upon, and at the same time, doing justice to people, and helping them feel represented, and seen in that data, and having it be a representation of their world. And that makes the data more complex in Wikidata sometimes. We regularly get told that our data is hard to use because of that. And if I could change one thing, it would be finding ways to bridge that gap.

Image by Whose Knowledge?, CC BY-SA 4.0, via Wikimedia Commons

Lydia Pintscher —

WATCH THE RECORDING OF THIS PANEL ▶

Imaginations and Implementations

During this part of the event, participants split up into seven small groups that combined members of different backgrounds and expertises. Each group had a facilitator, as well as someone in charge of listening and reporting back to the larger plenary. This time, we posed questions aimed at the future: what would structured data with multiple epistemic frames and anchors look like? We asked folks to imagine an internet in 2040 that has structured data with multiple epistemic frames and anchors at its core.

Groups detailed different elements related to the topic: the need to recognize that lived experiences go far beyond online (structured) data; the urgency to create, promote and expand initiatives around literacy and education about data and its impacts on people, and the possibilities brought up by centering marginalized communities in the design, ownership, and leadership of online platforms.

See snapshots from these conversations:

structured data that can be built without having copyright problems around databases, moving away from the copyright/licensing and + about ethics of data

Data archives and datasets that allow for modes of contestation -- whether that involves tinkering, challenging, or outright refusal.

Diversificação das equipes: garantir a presença de diferentes vozes e diferentes áreas do saber para criar espaços de criação mais ricos e diversos.

Menos centralização - mais diversidade de plataformas e opções

Plataformas open source - para permitir a alteração dessas infraestruturas que sustentam as plataformas

Plataformas open source - em que podemos acessar as infraestruturas e entender qual a base de conhecimento que as sustenta.

an env. sustainable way to collect / process / distribute the data - not a nice thing "you're now on the Internet but now I'll take your water with my huge data center!"

Structured data that invites challenge and rejection as a core affordance of the data and its representation.

There is the need for us to create conversations around diversity and digital divide when it comes to knowledge-sharing. Not many people can understand the most visible challenges.

Clarity and transparency on what a structured data system can and cannot depict and interpret

Can it also be assumed that these multiple epistomologies will be visible to all? I think that Indigenous communities will not necessarily make their knowledge visible

internet is that people/users are expected to contribute to a decolonising when there is no transparency around the structure itself i.e. how can you dismantle something you don't understand

this is also related to corporations keeping things secret. Open Science Policies vs Major Corporations Practices. The influence of research funding on data modelling and data collection.

The ability to re-write ones own past data (in archives) to map ones own development and growth.

Variations in access and governance to the data, depending on whose knowledge it is and who should have the ability to view and see it.

Representation of communities in the ownership/leadership, design, content, editors and in the downstream applications of the data

structured data are sensitive to our local knowledge embedded in songs, tales, etc.

better processes for ethically collecting and stewarding the data -- less power dynamics and more collaboration dynamics

structured data easy to understand, not only for experts but for everybody

creating safe space for discussion

The 'internet' itself is a structure people participate in but which very few people actually understand. How to reconcile the two, especially within a context of wanting to decolonise?

Using a Fanon-esque framing of goal of decolonisation to "change the order of the world" what potential does structured data have to reflect those new orders?

how can structured data help us to "disorder" our world?

sustainably funded institutions with explicit goals of reparative justice, decolonial approaches, indigenity, emancipatory and liberatory practise

Logical constraints and data models are keys to decolonisize structured data.

Smaller sovereign data sets/taxonomies not BIG data??

Structured data are systems of description of the world. All description is interpretation, at least insofar as it depends on "terms", and terms are distinctions that are frequently culture-specific and non-universal.

Is the solution in the ontology, vocab, standards? or is there another place? What could it be? Conversations between users?

There could be a use for fugitivity for the ppl that are working to subvert the institutions and also the possibility of marginalized communities to make use of that as well.

subversion and sabotage towards to goals of liberation and emanipation? Knowledge justice / debt

Thinking about the idea of complicity (inc mine) for the people who are working at the hegemonic structures

Defining principles instead of ontology (IEEE indigenous data)

Structured data in multiple languages, not only the principals, considering the diferentes concepts or ideas, considering gender, ethnic groups and people with vision or auditive problems

Pensar los datos en las realidades diversas que los contruyen, las relaciones de poder presentes en ese proceso, cómo transformar y abrir dicho proceso, hacerlo inlusivo

como pensar la diversidad de culturas y territorios. Entender que no todo se puede traducir y que las realidades exceden los datos en si. Imaginar un mundo de datos estructurados contextualizados.

Si lo pensamos desde una perspectiva en la que el ciudadano y no solo los estados puedan realizar estos datos estructurados podría ser más educativo y beneficioso para reflejar la realidad que se vive

Happy to participate in the Breakout Rooms of @WhoseKnowledge Decolonizing the Internet's Structured Data. The most important outcome here is to think that seeking fairness in #KnowledgeGraphs should not only be restricted to facts but also to data models and logical constraints.

HOUCEMEDDINE TURKI
@Csisc1994 on Twitter, 13 October 2021

Listening and Learning

This session called *Listening and Learning*, followed our smaller group discussions. We invited listeners from each group to report back to the plenary and, once again, we had an amazing group of members: Thomas Hervé Mboa Nkoudou, a Cameroonian social scientist, biochemist and maker; Maya Indira Ganesh, an Indian researcher and educator on tech, especially ethical AI; Jani Pereira, a biologist based in São Paulo, Brazil; Constanza Verón, a historian and educator based in Buenos Aires, who is currently at Wikimedia Argentina; Asaf Bartov, a Wikimedian, structured data enthusiast, a digital librarian; Amanda Jurno, a Brazilian professor and researcher in communication studies, from Wiki Movimento Brasil, and Alex Hanna, a sociologist and Director of Research at the Distributed AI Research Institute.

They explained what had surprised, excited, or challenged them and what had been the most radically imaginative and interesting ideas — as well as suggestions for how to get there. Our

listeners and other participants offered these key insights:

🏳 ACCESS AND CONTROL OF DATA

▸ Recognize that decolonizing structured data and the systems built upon them is a critical and urgent human rights issue, not simply a nice-to-have conversation on the margins.

▸ Acknowledge that marginalized communities — the minoritized majority of the world — have little to no access or control over the data that governs their lives currently.

▸ Share knowledge with different communities about data, technical tools, and how to inhabit and use online spaces and digital tools safely and fully.

▸ Be aware that a community-centered vision for structured data may be irreconcilable with the agendas of some individuals and organizations invested in structured data, especially those using it for profit over human well-being.

🏳 AGENCY/ENGAGEMENT

▸ Deny people the excuses not to engage with the decolonization of structured data, by creating and sharing resources and gradually create an ecosystem where there are no reasonable excuses not to acknowledge, and represent diversity of contexts and epistemologies

▸ Acknowledge that Indigenous and other marginalized communities should have the right to refuse to have their knowledges online, or to be datafied, and the agency over what information they share with the world.

DISTRIBUTED DATA

▶ Develop ways of creating smaller and connected datasets that are governed by marginalized communities themselves, rather than by companies based only on profit.

PLANET-CENTERED REDESIGN

▶ Be mindful of and incorporate in discussions the environmental impacts of the infrastructure needed for the processing of massive amounts of data (e.g., data centers).

▶ Restructure systems while constantly asking ourselves: what are they for? What purpose do they serve?

PLURALITY OF DATA

▶ Redesign systems and model them based on different knowledges and communities' glorious complexities.

▶ Recognize that most of the world's knowledge systems are not text-based (and might include a medium that involves songs, tales, artworks of different kinds), and go beyond text to find different forms of encoding data.

▶ Learn from local examples and their specificities in order to create broader changes, especially because "local" is rarely "small", especially in the Global South.

▶ Commit to listening and accommodating different perspectives in changing the nature of structured data, knowing that decolonization is a process that takes time and constant effort from multiple people and communities.

What is next?

For many participants, *Decolonizing the Internet's Structured Data* was a much-needed opportunity to pause for a moment and imagine radical possibilities for structured data. It also meant a chance of doing it in community and across regions and areas of study and of work. Moving forward, many of them voiced their interest in organizing more collective spaces like this and advancing more concrete steps towards emancipatory practices.

As organizers, we recognize the enormity of the challenge, and we knew from the start that this conversation would need to be the first of many. As we move into 2022, we are keen to create and convene more opportunities to radically reimagine and redesign the internet's structured data through a feminist, anti-colonial, anti-racist lens. With these next steps in mind, we are staying connected, through a mailing list, with everyone who joined us for this first conversation. We will also keep our broader communities informed and involved, through updates in our newsletter.

Gratitude

Decolonizing the Internet's Structured Data would not have been possible without everyone who joined the conversation, thank you to all our amazing panelists, listeners and participants. Thank you to our fabulous interpreters as well, without whom this would not have been a multilingual conversation. We also want to express our gratitude to our co-conspirators for this event, Amanda Jurno and Érica Azzellini from Wiki Movimento Brasil, and Sabine Müller and Dominik Scholl at Wikimedia Deutschland. Last but not least, we take this opportunity to thank our incredible supporters, partners, and allies at large, whose everyday support made this conversation happen, and will make the many more to come a reality.