Deep Diving into Decolonizing Structured Data

Summary report

In August 2023, over 30 participants from more than 15 countries were brought together for lively and insightful conversations on decolonizing the Internet's structured data. The sessions took place online on 8th August, and in-person on 15th August. The discussions were held ahead of Wikimania 2023 in Singapore, and were organized by Whose Knowledge? as a continuation of the Decolonizing the Internet's Structured Data event in October 2021. This report is a snapshot of the provocations, reflections, and happy speculations that emerged from our collective thinking about structured data.

Image generated by OpenAI, 2023. Prompt: Illustration of structured data, risograph

Contents

Why this gathering?

Structured data is at the core of how the Internet — as we currently know it — works. They are pieces of information organized in such a way that they can be easily read, understood, and processed by machines. Massive amounts of data get sorted, organized, classified and linked to other pieces of data. Like any piece of knowledge, these data points are informed and structured by and around specific regulations, traditions, and epistemologies.

Across a series of **Decolonizing the Internet convenings**, we have had ongoing conversations about issues of power and privilege inherent in the ways knowledge is understood and the ways the internet is designed and experienced, in particular for/by/ with women, non-binary, racialized peoples, LGBTQIA+ communities, Indigenous peoples and others from the Global Majority. With the 2021's Decolonizing the Internet's Structured Data gathering, we convened with the intention to deepen the discussions, by analyzing the power dynamics of structured data, and whose views, whose agenda, whose ontologies, and whose decisions build and sustain these classifications and systems.

These conversations have evolved in many different contexts of our work and are deeply interwoven. Among them are: #VisibleWikiWomen, Language Justice and Liberatory Archives and Memory.

Overall, we are concerned about how massive processes of colonial-capitalist datafication are reinforcing structures of power and privilege – the rarely discussed backdrop of AI hype.[1]

This year's **Learning Circle: Deep Dive into Decolonizing Structured Data** was held as a continuation of conversations started at 2021's Decolonizing the Internet's Structured Data event. Over two separate days, we held space for activists, artists, technologists, scholars, linguists, librarians, wikimedians, academics, feminists and many others to have conversations.

As a preamble to Wikimania 2023, it was our intention to create an intimate space with friends and allies with different entry points to structured data, to collectively learn together, and build a shared understanding for next actions in advancing more concrete steps towards data emancipatory practices.

That 'who' in "who is making the questions?" is more a mindset than people. Which […] mindset is behind taxonomies? (Corporation mindsets are a problem for diversity, but not only. We do have this trouble in Wikimedia projects, as well).

– Vic Sfriso

Image by Radhika Bhardwarj - Taken during an event, CC BY-SA 4.0 via Wikimedia Commons

What we did and how we did it

Learning Circle: Deep Dive into Decolonizing Structured Data was held over two sessions. We held our first Learning Circle online on 8 August 2023, as a way to guide the conversation collaboratively. This was followed by an in-person gathering on 15 August. Each space was guided by our learning principles of love, respect, and solidarity. This was a liberatory experience, informed by bell hooks' feminist pedagogy (Teaching to Transgress: Education as the Practice of Freedom) with space for both individual and collective multidirectional learning, recognizing our diverse points of departure and arrival, our own power and agency, and our different paths.

The Learning Circle was a collective learning process for conversation, questions, exchange of ideas, and flow around experiences, where participants arrived from their knowledge centers. The Learning Circle methodology is inspired and derived from talking circles or circle work used by some Indigenous communities to also share, listen, explore and embody knowledge. This method lays weight on "a structure where the voice of each person present is given equal weight".[2]
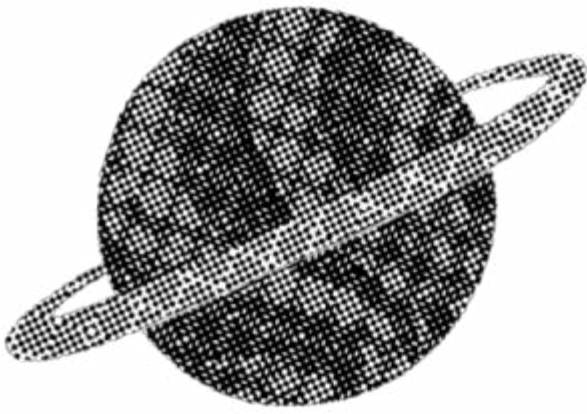
Remote Learning Circle

The online space was loosely facilitated by Mariana Fossatti, Decolonizing Wikimedia coordinator, and Sunshine Fionah Komusana, #VisibleWikiWomen coordinator. They presented the zine that was inspired by 2021's Decolonizing the Internet's Structured Data report. As part of accessibility approaches, each page of the zine was verbally described to accommodate the visually-impaired in the room.

Visual description

Deep-etch image of two South Asian femmes who are looking down, slightly smiling and digging into soil. They are overlaid on a black and white image of European and North American currency bills.

Visual description

```
Half-tone risograph image of the
planet Saturn.
```

In conceptualizing the Learning Circle, we wanted to have a space for people with different skills and interests in structured data to talk to each other. To explore this, we invited insight offerer, **Ishan Chakraborty**, from Jadavpur University to begin the conversation by narrating his experiences and challenges navigating websites and tools built upon structured data as a visually-impaired person. Additional insight offerers then shared their prompts and provocations about structured data, and to provide more urgent questions for the collective to think about.

George Oates, co-founder and Executive Director of Flickr Foundation; **Maari Maitreyi**, Knowledge Justice Researcher of Whose Knowledge?; **Flavia Doria** of Wiki Editoras Lx and Art+Feminism; and **Asmelash Teka Hadgu**, co-founder and Chief Technology Officer of Lesan further explored structured data issues together.

The format of this conversation was a free-flow talk among the group, around the question: **What are the relevant / interesting / urgent**

Visual description

```
Half-tone risograph image of a cinema
clapperboard and half-tone risograph image of a
cellphone with a map navigation system active.
```

questions you can offer to this collective to think about decolonizing structured data? Our intention was to bring together not only the expertise of this particular group of insight offerers, but also their own unanswered questions that could be further addressed in the Learning Circle.
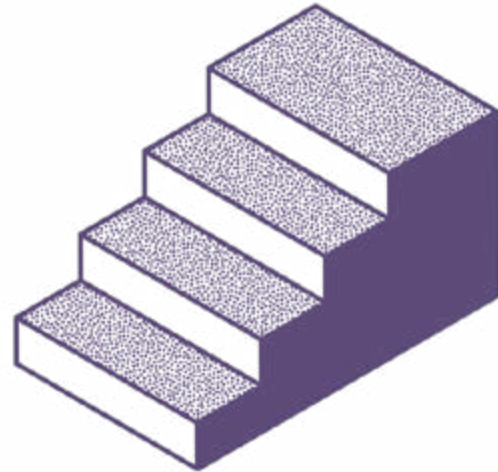
Visual description

```
Risograph image of a staircase in dark
purple.
```

From here, participants entered breakout rooms according to their language of choice (English, Spanish, and Portuguese) to answer the question: "**Based on what you have heard today and from your experience, what is one thing you want to make sure is discussed in Singapore, and beyond?**" We used the responses to set the agenda for the in-person Learning Circle in Singapore.
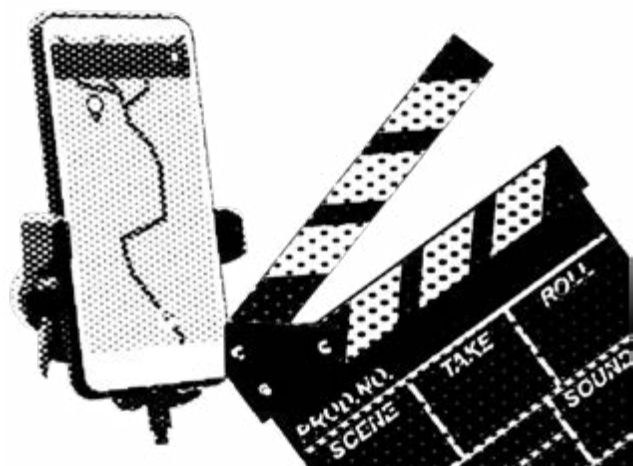
In-person Learning Circle

In Singapore, we welcomed 16 participants to our in-person Learning Circle on 15th August to dive deeper into the topic of decolonizing structured data. To set the tone, participants created their own human-made knowledge graph[3] to map, with data, who is in the room. In the activity that followed, they shared observations and reflections from the exercise and identified insights that stood out from the online gathering. Then, they had the chance to choose one case study on structured data to learn from. After getting into groups based on these case studies, they shared their findings in a circle together. In the final part of the day, we held space for each person to come up with "happy speculations"[4] — a liberatory action towards decolonizing structured data, an action to take forward — for the future of structured data.
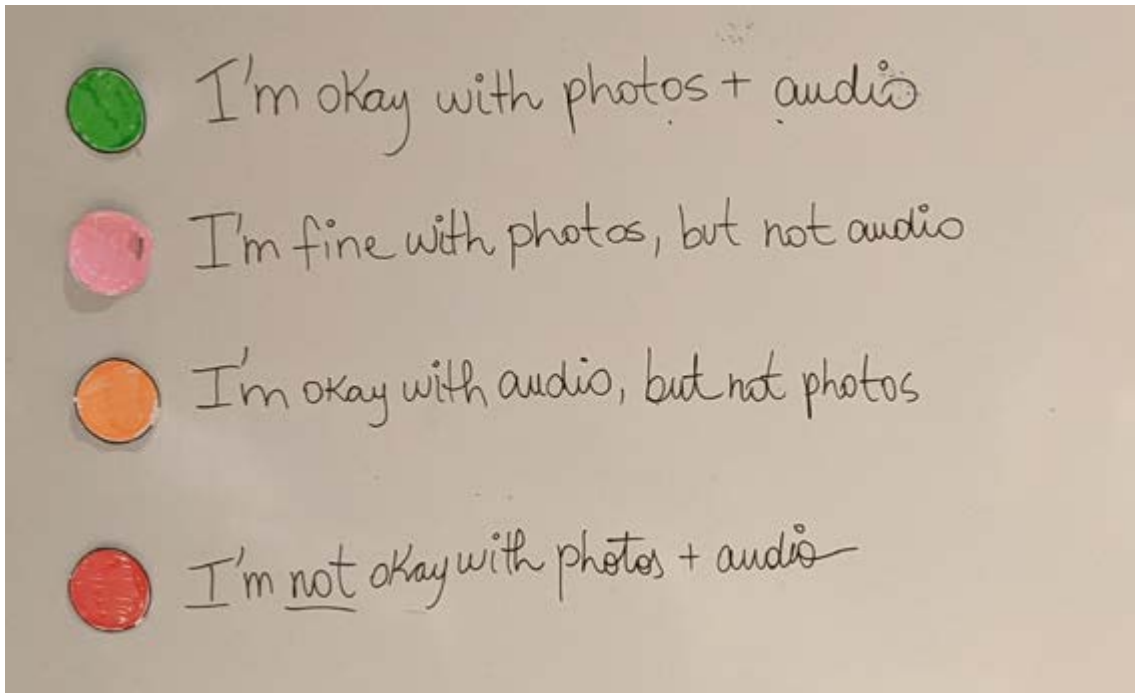
Left and below: Image by By Radhika Bhardwarj - Taken during an event, CC BY-SA 4.0 via Wikimedia Commons

For the in-person gathering, in particular, we placed emphasis on consent to ensure everyone's privacy and safety. Participants had the option of using four coloured stickers (green, pink, orange, and red), based on whether they consented to have their pictures and audio taken and shared publicly, taken but not publicly shared, or not taken at all. We consider consent as something that cannot simply be assumed, but rather affirmed, constantly negotiated and made as an informed decision in relation to a person's context and surroundings. Both Learning Circles included health breaks for attendees to recuperate and be at ease whenever they found the sessions overwhelming or needed space to process.

WK? at WIKIMANIA SINGAPORE

In many ways, these Learning Circles were a callback to our first Decolonizing the Internet gathering in 2018, Cape Town, and the Decolonizing the Internet's Languages meeting in 2019, London. Similarly to those convenings, we brought together allies, friends, and collaborators to strategize and think through ways of centering plural forms of knowledge on the internet, its designs and the structured data upon which it is built. This took place ahead of occupying Wikimania – the Wikimedia movement's annual conference, organized by volunteers and hosted by the Wikimedia Foundation.[5]

This year, Whose Knowledge? facilitated three events at Wikimania Singapore:

Decolonizing knowledge through languages: why making Wikimedia more multilingual and multimodal matters

Our Language Justice coordinator, Claudia Pozo, and language activists and wikipedians Dumisani Ndubane (South Africa), Ishan Chakraborty (India), and Paska Darmawan (Indonesia) explained their approaches and challenges with multilinguality in the wikiverse.

Image by YoulendreeAppasamy - Own work, CC BY-SA 4.0 via Wikimedia Commons

#VisibleWikiWomen Lab: fostering multilinguality and decolonizing structured data narratives on Wikimedia Commons

This in-person workshop and talk presented the #VisibleWikiWomen category on Wikimedia Commons as a site of experimentation with metadata, such as image descriptions, to tell a story. Our #VisibleWikiWomen coordinator, Sunshine Komusana, and Decolonizing Wikipedia coordinator, Mariana Fossatti, facilitated this workshop, guiding us through the potential of structured data on Commons.
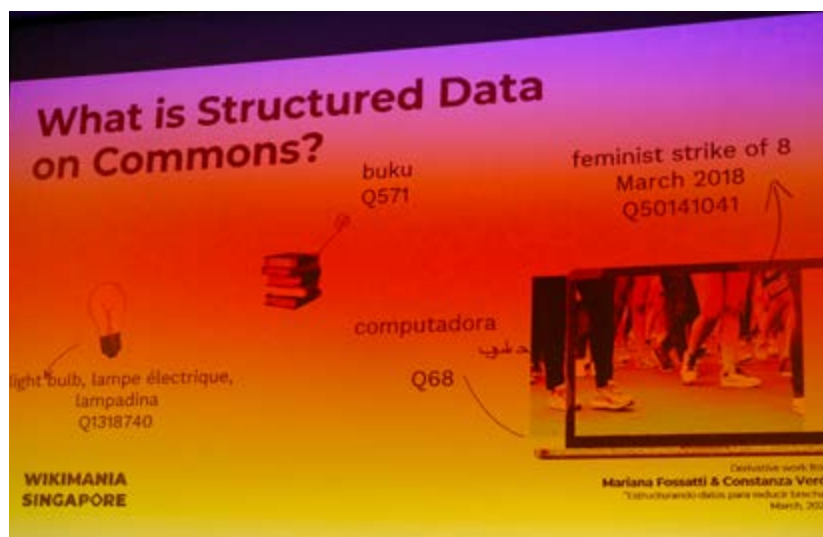
Image by YoulendreeAppasamy - Own work, CC BY-SA 4.0 via Wikimedia Commons

#VisibleWikiWomen photobooth

Images by YoulendreeAppasamy - Own work, CC BY-SA 4.0 via Wikimedia Commons

The photo booth was conceived as a physical space for image-taking, and as a feminist corner in big convenings like Wikimania. It made memories tangible through documenting women, non-binary people and feminist allies at Wikimania and posting these images on Wikimedia Commons, making the platform more illustrative of the majority of the internet's users.

Image by YoulendreeAppasamy - Own work, CC BY-SA 4.0 via Wikimedia Commons

What came up?

Remote Learning Circle
What's one entry point to structured data in your work?

Want to know more about descriptors for digital objects

Curious how the non-Anglophone structured data work

Open GLAM

Programming and coding

Wikipedia infoboxes

Wikidata through Pywikibot

Folksonomic/organic digital information system

Interested in structured data and image descriptions

Wikimedia projects in culture and heritage

Wikimedia projects e.g. Wikidata and Commons

Indonesia Wikipedia and Wikidata

In a shared workspace, participants explained their primary entry point to structured data in their work, and their curiosities about structured data.

What are we missing when we talk about accessibility of structured data?

Ishan Chakraborty prompted the insight offerers with a valuable contribution about his lived experiences online, as a visually impaired person whose language of choice is Bengali. Many websites take very little consideration to accessibility needs — from languages available to platforms that are "almost impossible to navigate," he explained. As such, this can affect accessibility to knowledge, as only able-bodied folks and those who speak regionally dominant and/or colonial languages, such as English,

can access information faster and easier. Disabled and visually impaired persons also lose out on many opportunities and information as a result of this inaccessibility.

The accessibility issues also add concerns around privacy and confidentiality, as he had to depend on others to assist him to access these sites.

Most websites use some form of structured data. They are not only inaccessible to persons with visual disabilities, but also limited when it comes to language. I would prefer to access this information in my language Bengali, but they are mostly in English. [...] So people like me remain out of loop and miss out on opportunities due to this.

— Ishan Chakraborty

Since I have to depend on others to access these websites at work, such as with the help of coworkers or a caregiver, are there ways to bridge the confidentiality gap?

— Ishan Chakraborty

Marginality Within Marginality

To know more about Ishan and his work, you can read an article he penned for our State of the Internet's Languages Report, available in Bengali and English.

Offerings to the Learning Circle
One foot on the land, one foot in the sea

கடலில் ஒரு கால்,
கரையில் ஒரு கால்

Conversations about structured data, more so than not, invite more questions than solutions. The challenge: which questions should we prioritize when it comes to decolonizing structured data?

Insight offerers reflected on the challenges of keeping up with tech advancements and surviving within existing data practices that were rooted in many forms of oppressions, while at the same appreciating what George called "the variety and poetry" [un]structured data points and descriptions could offer.

An ongoing metaphor was provided by Maari, who paraphrased a widely circulated Tamil proverb about having one foot on land, and another in the sea – Kadalil oru kaal, karaiyil oru kaal ["கடலில் ஒரு கால், கரையில் ஒரு கால்"]. This was her reflection on feeling the pressure to deal with an increasingly AI-saturated world seemingly overnight and feeling the need to simultaneously pause to connect and check in with communities. Is representation the best we can hope for?

DATABASES HAVE HISTORICALLY EXISTED WITHIN THE ETHER OF HISTORY OF COLONIALISM, CAPITALISM, PATRIARCHY, AND OTHERS. AS WE USHER INTO THE NEW TECHNOLOGICAL ERA OF MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, AND AUTOMATION, IT FEELS LIKE MANY OF US ARE THROWN INTO THESE BINARY MODALITIES OF SURVIVAL AND ENGAGE WITH THE TECHNOLOGY OR IGNORE, BE INVISIBLE AND PERISH. HOW DO WE HOLD ON AND NOT YIELD TO THESE PRESSURES AND MAKE OUR OWN PATHWAYS FORWARD?

– MAARI MAITREYI

I AM A BIG FAN OF ORGANIC DESCRIPTION, ESPECIALLY WHEN IT COMES TO IMAGES — SOMETHING COMPUTERS CAN NEVER OFFER. HOWEVER FRUSTRATED WE MAY BE BY THE NEED FOR COMPUTERS TO DO THESE GIGANTIC [MACHINE] CALCULATIONS FOR THE DESCRIPTIONS, WE CAN, AT THE SAME TIME, APPRECIATE THE 'VARIETY AND POETRY' IN THEM.
— GEORGE OATES

Facing a massive amount of data these days, we also spoke on how to meaningfully organize and preserve the knowledge. We also thought of possibilities of different ways to deal with memory-making and archiving in a way that does not follow the hegemonic extractive Western culture.

IT'S IMPORTANT TO REMEMBER THE IDEA OF 'THE MORE DATA, THE BETTER' COMES FROM A VERY WHITE WESTERN EUROPEAN CULTURE. HOW DO WE LEARN FROM INDIGENOUS PEOPLE ABOUT THE RIGHTS TO BE FORGOTTEN WHEN IT COMES TO MAKING AND PRESERVING MEMORIES? WHAT ARE OTHER PERSPECTIVES WE CAN LEARN FROM?
— FLAVIA DORIA

We also spoke about the limitations of open source and its relations to unpaid labor, the possibility of community-centered tools and practices, and demanding accountability from corporations benefiting from the work we do in structured data.

I PERSONALLY LOVE OPEN SOURCE — THAT'S WHERE I LEARNED MOST THINGS. BUT WHEN BIG TECH SCRAPED THE OPEN SOURCE DATA WE GATHERED AND PROFITED FROM IT FOR THEIR OWN TECHNOLOGIES, THERE'S NOT MUCH WE CAN DO [AS SMALL PLAYERS] WHO RELY ON WHATEVER TINY RESOURCES WE CAN TO CURATE DATASETS AND BUILD TECHNOLOGY FOR OUR LANGUAGES AND COMMUNITIES. HOW DO WE ALSO EMPOWER GRASSROOTS COMMUNITIES TO BUILD THEIR OWN TECHNOLOGIES?
— ASMELASH TEKU HADGU

How do we contend with the openness of open source?

For open source tool builders – what happens when Big tech lifts the datasets gathered to profit from it, and causes harm to the communities who inputted their data? People put a lot of unpaid labor into building up open source data or content repositories. This work is then taken for free by Big Tech corporations who in turn close off and make a profit from it. How do we deal with these unintentional outcomes of open source processes?

Revisit and deepen ethical data sharing practices in open source, towards taking care of our communities.

How do we challenge the existing status quo? How do we take back power?

How do we keep up with technological advancement forwarded by Big Tech e.g. machine learning, AI? How do we not succumb to this?

How do we demand accountability from these powerful parties benefiting from open source data?

How to re-examine the practices of structured data through a community-centered lens?

How do we bridge the people in tech and companies who are building our tools, with the needs of our communities?

What can we learn from Indigenous communities about autonomy, agency and the "right to opacity", as Edouard Glissant calls it? For example, learning from Indigenous communities and their rights to be forgotten.

What historically-rooted practices can we return to in order to re-examine our relationship with data?

How can communities empower themselves to build their own technologies suited for their needs, desires and capacities?

If tech and tool builders are not a part of communities that they build tools for, what are the steps to seek consultation from the communities before creating any tools?

What other ways do we navigate structured data amidst the sea of data?

How do we foresee the repercussions in the work we do?

How do we see the unintended consequences out of the structured data we organized and the tools we built?

How do we create useful and meaningful data structures without disembodying its contexts?

What do we need to have *complex* and *liberatory* conversations about structured data?

How do we design for people in the margins, first and foremost?

How can structured data reflect the complexities and realities of our communities?

LET'S NOT CONTORT OUR HUMANITY FOR COMPUTERS.

— GEORGE OATES

What does a model of refusal and resistance look like?

How do we design for people in the margins, first and foremost?

How to centralize accessibility in many forms of structured data? How to advocate for the plurality of lived experiences of people who exist at the intersections of multiple forms of oppression?

How do we design for users who lack meaningful connectivity[6] to the internet? In many parts of the world, people still depend on mobile phones to access the internet.

How do we make learning tools accessible for newly structured data literate users? In an environment where editing even the relatively accessible projects like Wikidata require skill-building, how do we ensure active participation in various structured data spaces?

How can structured data reflect the complexities and realities of our communities?

What do data practices look like when it comes to queer communities, representing multiple gender identities across the spectrum?

How do we include oral information that is not documented in written manner, especially when we consider that many Global Majority cultures are oral cultures?

How do we accommodate under-resourced communities who remain unsupported by tech? For instance, Indigenous languages from the Global Majority might only get access to information translated to their languages much more slowly in comparison to colonial or regionally hegemonic languages.

How do we recognise and support the — often unpaid — labor of people and communities who maintain the works of structured data?

What does a model of refusal and resistance look like?

Faced with massive data and the advancement of tech and machine learning by Big Tech, how do we advocate for data justice for our communities?

How do we take space in data governance, and how can we govern our own data on our own terms?

How do we decolonise structured data on the internet, when we do not yet have everything we need to represent all our needs and differences?

Image by Radhika Bhardwarj - Taken during an event, CC BY-SA 4.0 via Wikimedia Commons

In-person Learning Circle

Knowing each other through structured data: creating our own human-made knowledge graph

In this session led by Mariana Fossatti, participants took part in a human-made knowledge graph exercise. On sticky-notes, they filled in their names, the languages they speak, where they come from, and which organization(s) they are affiliated with. Using the same sticky-notes, threads and sticky tape, they then laid out the connections between each other on a table.[7]

Listen to this audio snapshot of Mariana explaining the exercise

Participants offered their observations as they worked on the exercise. Everyone in the room spoke English. A couple of attendees added layers to the family of languages they speak in — for instance, Multicultural London English (MLE) or Brazilian Portuguese — showing the complexity of data to represent our multiplicities and complexities even within a seemingly neutral category. From this basic piece of information, we began to expand our questions. For instance, in our exercise there was no way to infer the level of fluency for languages we speak, as well as who makes the decision as to our level of fluency. We also spoke about the ISO language codes[8] in structured data, and how it does not accommodate many other vernacular languages.

Listen to this audio snapshot of Kelly, Ilham and Alice speaking about ISO language codes

We also questioned decisions that are central to many languages and knowledges being sidelined on the internet. We also questioned Mariana's authority and decisions on the design of this exercise! Participants also learned that while the information on the sticky-notes on the table can imply many connections, they cannot infer information that comes from other sources. For example, how the languages a person speaks are related to the colonial histories of their locations. This is the same with machines and computers, which can only process data inputted into them. As such, in order to interpret structured data, we need to juxtapose it with other sources for more contexts.

WHO MAKES THESE DECISIONS? WHAT EFFECT DOES IT HAVE ON THE VISIBILITY OF CULTURES? [...] WHEN PEOPLE MAKE THESE DECISIONS 'AS NOT TO COMPLICATE THINGS', WHAT TRADE OFF DO OTHER COMMUNITIES HAVE TO MAKE?

– MAARI MAITREYI

Remote learning circle to in-person – what did we prioritize?

A LOT OF STRUCTURED DATA IS NOT MULTILINGUAL BY DESIGN, BECAUSE HUMANS THINK IT IS CONVENIENT TO HAVE ONE DOMINANT LANGUAGE.
– KELLY FOSTER

IN THE BIG DATA REALM, WE ARE TAUGHT THAT IT IS BETTER TO HAVE MORE DATA AND BE AS SPECIFIC AS POSSIBLE. BUT WE ALSO FAIL TO DISCUSS HOW PUTTING AS MUCH DATA AS POSSIBLE ONLINE CAN OPEN UP POSSIBILITIES FOR SECURITY ISSUES. WHAT IS THE LEVEL OF OPENNESS WE SHOULD HAVE FOR STRUCTURED DATA?
– ILHAM NURWANSAH

FOR DECOLONIZED INTERSECTIONAL STRUCTURED DATA TO EXIST, WE NEED TO THINK ABOUT DESIGNS THAT STARTS FROM THE MARGINS – FROM THE PEOPLE WHO ARE FURTHEST FROM THE CENTRE, WHO ARE THE MAJORITY OF US IN THE WORLD, THAT IS HOW WE DISRUPT THE POWER OF THE PEOPLE AT THE CENTER OF THE CURRENT TECH AND DATA SYSTEMS.
– SUNSHINE FIONAH KOMUSANA

WHEN WE THINK OF ACCESSIBILITY FOR VISUALLY IMPAIRED PERSONS, ISHAN MENTIONED [IN THE ONLINE SESSION] THAT THERE WERE NO FUNCTIONAL SCREEN READERS THAT ARE OPEN SOURCE. MANY HAVE TO USE PROPRIETARY SOFTWARE THAT IS NOT ALWAYS ACCESSIBLE [BECAUSE THEY NEED TO PAY THE CORPORATIONS DEVELOPING IT]. DISCUSSIONS AROUND ACCESSIBILITY TOOLS IN OPEN SOURCE ARE STILL LACKING.
– JORGE GEMETTO

Learning from case studies

The project **WikiData Project Chattel Enslavement and Freedom** aims to improve how data about enslaved people, self-liberated people, and data of the histories of racialized chattel slavery is modelised and structured; add knowledge created by enslaved people; and develop ethical approaches to data modeling about enslaved, enslavement, and freedom. This project seeks to answer the primary question about the project: "What is the impact of labeling and "datafy"-ing an entire group of people as property?"

In the group, we discussed the incredibly complex and nuanced issues of how to approach this with care, understanding, and centering communities who experienced the harm. There is a need to reimagine new ways of representation and framework to deal with historical records and contextualization. This is in addition to how Wikidata should be informed by the corpus of work addressing the ethics of digital humanities projects that study enslavement, taking places both within academia and by vocational-practitioners.

Listen to an audio snapshot of some of the histories of "enslaved people as property".

RACIALISATION IS CLASSIFICATION, AND IN ACTIVE CLASSIFICATION, YOU CAN RACIALISE SOMEONE.

— ALICE KIBOMBO

The **VisibleWikiWomen campaign** aims to create more visibility to women and non-binary people on Wikipedia and the broader Internet through uploading images to Commons, the media repository for Wikipedia and sister projects. Since the introduction of [Structured Data on Commons](), editors can tag images to connect them to Wikidata. [Computer-aided tagging]() is available for getting suggested labels, but this tool has not been free of problems and biases. In the group, we mulled about the careless representation and biases encoded in this automatic machine tagging. But, does this data tell the stories behind the images? What's at stake, when an image depicting a Black woman with a microphone is "seen" and tagged by the machine as a music artist but not as the politician or [filmmaker]() she actually is?

Sunshine explains more about this example.

Is this happening systematically to white women, to men, to cis people? Conversations around the politics of labeling and visibility came up, while at the same time being mindful of the biases fed into the computer — as well as the labor involved in correcting them.

The case study **LGBTT+ individual and collective representations in Wikidata** looks at representation of gender in Wikidata, looking at the way gender is currently modeled and how it had been done in the past. Current challenges include incorporating the fluidity of identities and name changes (current model has both sex and gender in the same taxonomy), especially in some places where gender identities can be tied culturally and geographically as a collective practice. In the group, we discussed how to build queer collective memories from structured data. We spoke about new ways of representing a wide group of collective gender identities in structured data. As a result, there were more questions than answers. For example, we asked ourselves: if we do not specifically label ourselves in a world in which one needs a certain identity to exist, how do we prove our existence?

Image by Radhika Bhardwarj - Taken during an event, CC BY-SA 4.0 via Wikimedia Commons

In the group **Wikilontar and Wikisource Loves Manuscripts data modeling**, we discussed the interpretation of structured data by non-wiki users through the examination of two cases: Balinese WikiLontar, a program to collect metadata and cataloging balinese palm-leaf manuscript in Bali to added more reference for Wikipedia, specially Wikidata; and Wikisource Loves Manuscripts, a project to digitize and transcribe over 20,000 pages of Indonesian manuscripts. The group had come up with some concrete solutions.

In an attempt to interpret data to non-wiki users, project lead Ilham Nurwansyah had crafted a plan to collaborate with Jorge Gemetto from Wikimedistas de Uruguay. The plan involved using the tool Jorge developed to visually represent data of cultural works, and make it easily searchable through the query field, and engaging for people to get involved in data editing.

Image by Radhika Bhardwarj - Taken during an event, CC BY-SA 4.0 via Wikimedia Commons

Happy speculations

Reimagining new ways of data structuring and justice

Identities, agency, and memory

Knowledge production and new projects

Access and knowledge

Make space, take space

Labour, resources, and maintenance

Reimagining new ways of data structuring and justice

Decide on new forms of data modeling to record the unrecorded — ones that do not view these groups as 'unknown' — and to think of decisions on whom to create these records.

Access and knowledge

Provide more access for informed contribution to fill in the existing gaps and the inconsistencies in structured data. This can also empower other groups to have more opportunities to be active participants in shaping the narrative and the direction of structured data.

Bridge the gaps and ensure access to structured data and justice, so that it is all easily available and accessible to people beyond our Learning Circle.

More collaborations to reduce gaps of knowledge in the Global Majority cultural heritage, and to model data in a way that makes sense to communities.

Remove barriers for the reimagining of structured data, especially centered on communities in the margins.

Identities, agency, and memory

More ways to incorporate the ever-changing and fluidity of identities, rather than reduce them to the fixed and static categories.

Make use of structured data to build collective memories and resistance.

Knowledge production and new projects

Produce works on "Black digital practice"[9] and the issues on the datafication of the trans-Atlantic slave trade for Wikidata.

Structured data project around Indian indentureship diaspora.

Collaborate to develop data visualization for global manuscript collections as part of humanity history access.

Labour, resources, and maintenance

More spaces, resources, and tools for corrective and gap-filling work that is more communal and focused on decolonial feminist perspectives.

Learn more about thingification and datafication.

Ensure continuity of the discussions we had into our everyday work.

Make space, take space

Have our own VisibleWikiWomen feminist Data Fest!

Instead of producing and collecting more data, we need a less quantitative approach to data. We need more plurality in data and more sensemaking for data justice. Instead of more datafication, we need more feminist data parties!

— Mariana Fossatti

Participant Reflections

Thanks to Whose Knowledge? for creating this space, facilitating this space, and holding the space for all of us. It's been really great. Something that I think about is sometimes we can be in these spaces and they can be so great, but then we'll have intentions of doing things afterwards. But then just the everyday work kind of takes over [...] so I'm really excited to talk about what we talked about here with my team, and see what we can do to help continue these conversations.
– Kira Wiesnieki

Because I'm a woman of few words, mine is really about informed contribution. A lot of the inconsistencies and the gaps that we see are because we've created a situation where people who have no knowledge of the history or knowledge of the subject, or who actually live these experiences, are contributing to this kind of thing. If you have informed contribution, it means that people who own this information, people who have generated it, people who are active participants in that experience, can use this opportunity to shape that narrative. And not only that, they can establish authority about how they want to be presented.
– Alice Kibombo

I'm going to connect the informed contributions and the spaces and resources, with hopefully bringing the data modeling days conversion for the data community, to come together around modeling complicated topics like we talked about today. And have a space to talk about it but also bring it to the attention of the community, that maybe some topics are more complicated than we currently make them.
– Lydia Pintscher

Here's a multilingual goodbye for now from Whose Knowledge? and in-person Learning Circle participants!

What's next?

Learning Circle: Deep Dive into Decolonizing Structured Data ended on a high note as we prepared ahead to participate in Wikimania 2023 in Singapore. Through a participatory, plural, safe, and caring process we managed to hold conversations and reimagine new ways of structured data practices that are not centered on Western, Global North hegemony. Conversations around communal consultation and inclusion, collaboration, labor, designing for plurality and not dominance, and representation cropped up many times in these two sessions. Finally, in the act of decolonizing, we also asked ourselves — what would a model of refusal look like?

We understand the work of decolonizing is ongoing and iterative. As such, we look forward to more collaboration, to more movement building and strengthening, to more plurality and beautiful mess, and to more joy and parties! In the meantime, we invite you also to check out the report for 2021's Decolonizing the Internet's Structured Data and to stay tuned for more updates on the Decolonizing the Internet convenings at Whose Knowledge? website, which includes a podcast series from these discussions.

Endnotes

1. O'Neil, Lorena. 2023. These Women Tried to Warn Us About AI. Rolling Stone.
2. Indigenous Knowledge Education
3. A digital structure that represents knowledge as concepts and the relationships between them. A knowledge graph can include an ontology that allows both humans and machines to understand and reason about its contents - https://en.wikipedia.org/wiki/Knowledge_graph
4. We credit the term to George Oates of Flickr Foundation, who attended our first Learning Circle on 8th August. The term is also inspired by feminist theorist Saidiya Hartman's idea of critical fabulation, who mentioned how design can work as a tool to "narrate a certain possibility".
5. https://en.wikipedia.org/wiki/Wikimania
6. Meaningful connectivity is when people can use the internet every day using an appropriate device with enough data and a fast connection.
7. This exercise was inspired by two great structured data pedagogical esources: Ashleigh Faith's videotutorial "Step-by-Step, No-Code Taxonomy Model ANYONE Can Learn" and "Cuadernillo sobre cómo enseñar con Wikidata", created by Wikimedia Chile.
8. ISO language codes are internationally recognized codes for more than 500 languages or language families. The most famous standard sets are ISO 639 (and the subsets ISO 639-1, ISO 639-2, ISO 639-3) for language codes, and ISO 15924 as standardization of language scripts. These standards also now play a crucial role in machine learning, which says a lot about exclusion of vernacular languages.
9. https://read.dukeupress.edu/social-text/article-abstract/36/4%20(137)/57/137032/Markup-BodiesBlack-Life-Studies-and-Slavery-Death

Our deep gratitude

To all the Learning Circle participants and insight offerers for bringing their multiple perspectives and selves to the conversations we had. A special note of appreciation to Mellon Foundation and Wikimedia Foundation, and our volunteers and collaborators for their energy and time in making these Learning Circles amazing. To many more spaces of co-conspiring and collective thinking!

Credits

Written by: Zana Fauzi with inputs from the Whose Knowledge? team and Learning Circle participants
Design by: Youlendree Appasamy
Art by: Youlendree Appasamy
License:
This report is issued under the Creative Commons Attribution-Non Commercial-Share Alike 4.0 International License (CC BY-NC-SA 4.0). In selecting this publication license, we welcome anyone to copy, excerpt, re-work, translate and re-use the text for any non-commercial purpose without seeking permission from the authors, so long as the new work is also issued under the same Creative Commons License.