Decolonizing the Internet's Languages

Summary Report, 2020

Decolonzing the
Internet's Languages
Summary Report, 2020

An initiative by
Whose Knowledge?

Written by Adele Vrana,
Anasuya Sengupta,
Claudia Pozo,
and Siko Bouterse,
with inputs from participants.

Layout and Design by
Claudia Pozo.

Lisenced CC BY-SA 4.0

Whose Knowledge?

DECOLONIZING
the Internet's Languages

Contents

Why did we meet? 6

Who was there? 9

What did we do? 11

What came up? 13

What's next? 27

Reflections 31

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

Why did we meet?

In October 2019, thirty participants from around the world were brought together by Whose Knowledge? just before MozFest in London, to scheme about Decolonizing the Internet's Languages. This is the report back from our gathering.

The internet we have today is not multilingual enough to reflect the full depth and breadth of humanity.

75% of the world's online population is currently from the Global South, and 45% of all women in the world are online. At the same time, we know that content online remains heavily skewed towards rich, Western countries, and most online knowledge today is accessible only through colonial languages.

We estimate that only about 500 of the world's 7000+ languages are represented online, with English and Chinese dominating. Google estimates that 129 million books have been published in about 480 languages. At best, then, only 7% of the world's 7000 languages are captured in published material. An even smaller fraction of these languages is represented in digital content.

When marginalized communities cannot create in their own languages on the internet, this reinforces and deepens inequalities that already exist offline. Most critically, those of us who are the primary consumers of digital content and infrastructure are still not

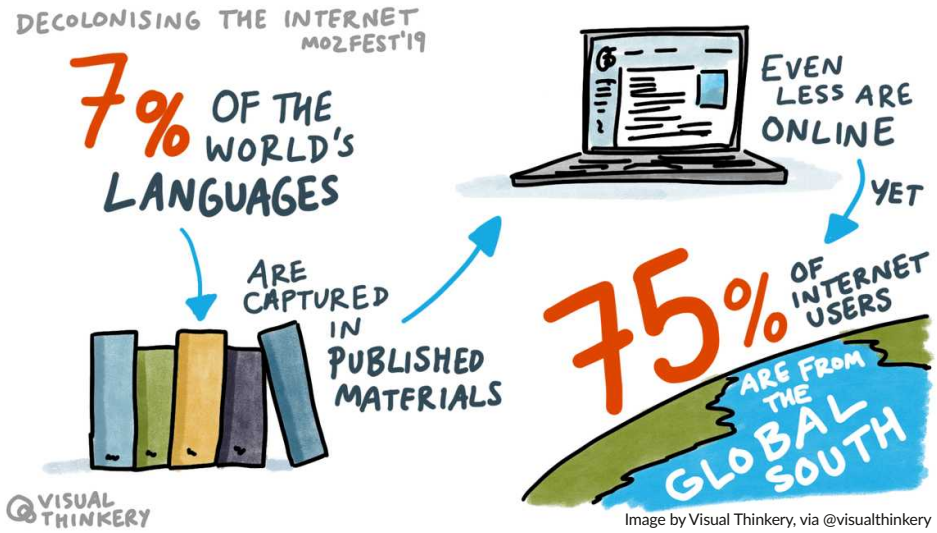


Image by Visual Thinkery, via @visualthinkery

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

the producers nor the decision-makers around its design, architecture, substance, and experience. The effort to change this – to re-imagine the internet and re-design digital knowledges – needs a multitude of us working together. So the idea for Decolonizing the Internet was born: an annual gathering, for unusual and unlikely allies to connect and scheme together towards a decolonized internet.

For this **2019 Decolonizing the Internet's Languages** gathering in London, UK, we brought together a diverse group of thoughtful, powerful folks who recognise that language is a proxy for knowledge, and who want to reclaim our many languages beyond English on the internet.

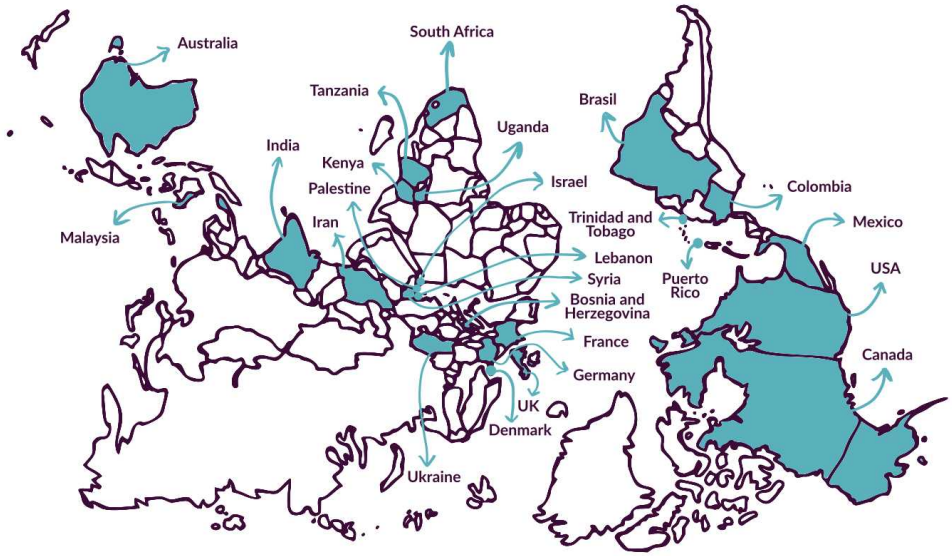
Our aim was to build alliances, and move from thinking together towards doing together - to take on various actions to decolonize the internet's languages, and encourage others to join us on this journey. Building a more multilingual internet is certainly a technical process, but even more critically, it is a social and political effort.

Who was there?

Participants at Decolonizing the Internet's Languages came from **more than 20 different countries**, bringing an incredible range of experience and expertise to the table. Rather than making this an open call conference, we invited each participant, aiming for a strong representation from marginalized communities while keeping this a safe and joyful space to hold these important conversations together.

A total of **30 artists, activists, archivists, mappers, translators, linguists, researchers, techies, funders, Wikimedians, Mozillans and much more**, participated in Decolonizing the Internet's Languages. 65% of participants were women or non-binary/trans folks, around half came from the Global South, and most spoke more than 1 language. This diversity in backgrounds and origins brought a wide range of perspectives and helped broaden the traditional frames for discussing the politics of knowledge, of language, and of the internet.

Decolonizing the Internet's Languages participants have roots in 24 different countries!



What did we do?

The Decolonizing the Internet's Languages gathering took place on October 23 and 24, two days prior to MozFest 2019 – Mozilla's annual conference – in London, UK. Our conversations were facilitated by Francesca Beard, who made sure we had a safe, inspiring and joyful time together.

Day 1:

We focused on getting to know each other and surfacing challenges and opportunities to decolonize the internet's languages. To kick off our convening, we spent the first morning listening and learning about each other's histories, stories and overall experiences bringing our languages online.

Our research partners, Sneha PP from Center for Internet and Society in India and Martin Dittus from Oxford Internet Institute in the UK, shared some preliminary data and next steps from our joint work to create an upcoming State of Internet's Languages report.

Moving from a place of deep respect, trust and solidarity, we began to surface the challenges and opportunities to create the multilingual internet we want and dream about.

Day 2:

We focused on building towards an agenda for action. Participants self-organised into groups in order to hold deeper conversations around the challenges and opportunities identified on day 1. Each group defined an issue or set of issues, brainstormed some potential actions, and thought about the processes and resources needed to support those actions.

You can read more about what we talked about in the section below.

What came up?

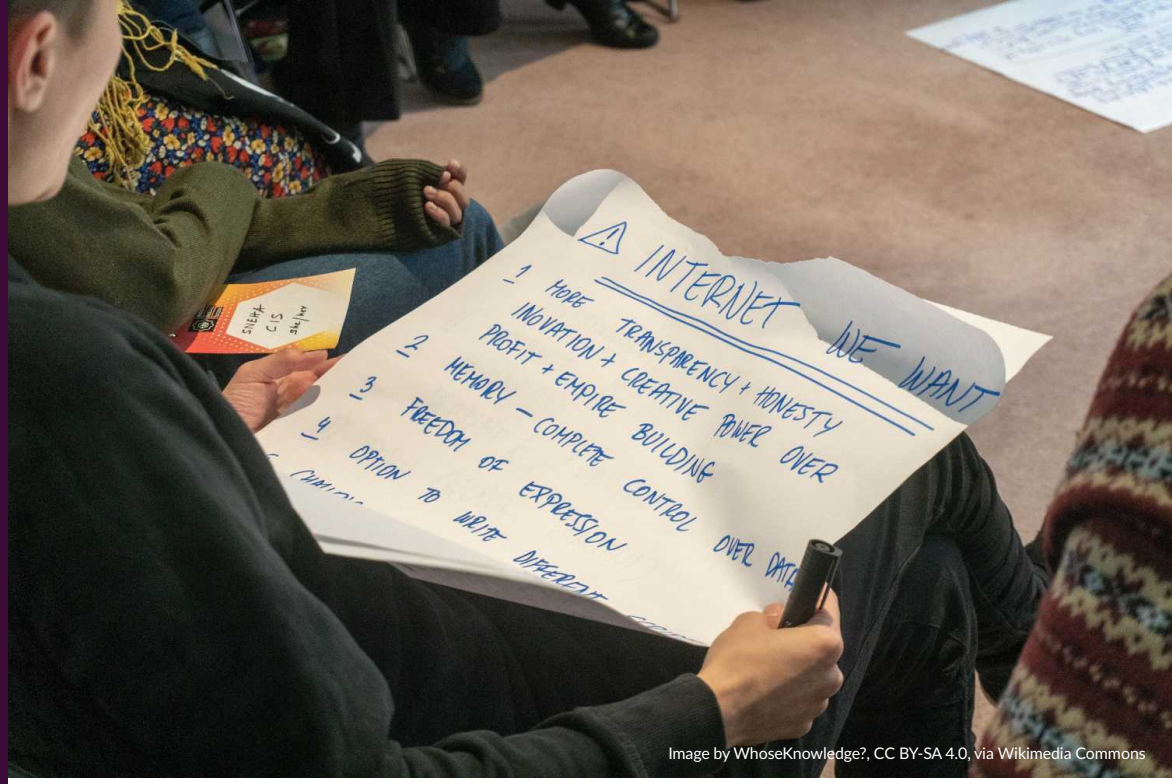


Image by WhoseKnowledge?, CC BY-SA 4.0, via Wikimedia Commons

We began by brainstorming some of the key challenges and opportunities to create the internet we truly want.

Challenges:

Capitalism and commercialization drive the internet, and are at the root of many problems with infrastructure, data, privacy, lack of sovereignty, bias, etc.

Profit is a motivator instead of collective good.

Internet, platforms, software, etc are mostly all proprietary and cost money (or personal data).
Materials for tech tools (phone or computer) are climate-damaging and lead to war and military interventions.
Individual mode of participation considers the user as a single agent.
The myth that Silicon Valley can lead a global tech/ governance structure.
The myth that open and free means safe and welcome. Many of our communities aren't safe online.
Many languages that are oral or don't have a script are excluded from the internet.
Local content discovery outside of search engines isn't defined in terms of public good.
Western epistemologies inform the internet as a space, with focus on technological and scientific determinism and the individual.
Security. For e.g., when there is a link between service providers and the state, the internet can be shut down, censored, etc.
Colonizers went out, recorded language, and told a story that wasn't theirs in archives/museums/etc. These are being digitized. What does this mean to the language and the knowledge on the internet?
Bias towards written content, whereas the majority of the world's languages are oral, some knowledge is visual, etc.
Access is not thought of as multidirectional. Communities who start to access the internet for the first time don't know how to search in their own language, and that language isn't represented on the internet until a community is also able to build its own content.

Opportunities:

Anti-capitalist centering of the internet, with anti-extractive and anti-oppressive principles.

Articulating a clear common set of ethics.

Collective/ common good as motivator.

Rethinking that the only way to access is through individual.

Centering the concept of sovereignty and looking through femninist and indigenious concepts of sovreignty and the embodied self.

Work in interdisciplinary teams.

Centering consent.

Combining tech and human curation to surface knowledge.

Privacy and the ownership of individual community data, including ability to remove it.

Community leadership.

Diversify tech builders, so that there is a plurality of structures and modes of engagement.

Decentralization.

Distributed networks that are local but globally connected.

Multiple formats - audio recordings, videos, writing, images, etc.

Public interest technology.

If code could not be copyrighted we'd have a very different internet.

The "minorities" who are actually the majority of the world should be at the decision-making table.

International agreements, such as through UNESCO, concerning internet languages.

Building a model for global governance.

Recognizing the importance of disability access - whether written or oral, all should be able to access and contribute.

Open source development.

Collective memory must have a seat at the table - just as important as the focus on “progress” and “future”.

Next we shared and clustered some key questions that could lead us towards potential actions for decolonizing the internet's languages.

Themes and Key Questions:

Documentation, Platforms, Participation

Language documentation is valuable, but costs money and required technical skills - how can we do it frugally yet effectively in our own communities?

Who is able vs. who is responsible for recording stories?

What measures for large marginalized languages can be used for smaller marginalized languages to prevent erasure?

How can we enable more African women to know about the possibilities of contributing to the internet's languages and to support them?

How do open platforms, discussions, projects, generate the participation for and by all the people?

What platforms exist for community documentation?

What education and training materials are necessary to understand the skills and challenges and dynamics in different contexts?

Research, Metrics, Advocacy

What does Google know that they could tell us?

What can or should we measure, and what shouldn't be measured?

How do we understand the challenge and change the way Unicode represents the languages of the internet?

How might we combine qualitative and quantitative data to do justice to the complexity of the embodied experience of the colonial internet?

How can we add nuance in the cause of language gaps?

How can we move Mozilla to advocate away from colonial languages? How do we do more advocacy work outside that framework and grow more diverse?

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

Orality

What are the other forms of communicating a language? What are the other ways of disseminating a language?

Can we think or imagine an oral Wikipedia, storytelling, having a platform for voices and video?

How do we consider affective registers, for e.g. Black Twitter, which has a unique language via gifs, emojis etc? How does Unicode take that into consideration, as a way a community talks and speaks about themselves?

How can we connect oral language speakers to the digital world? Both recording their own materials and accessing knowledge when they don't know keywords etc?

How can we document oral languages?

Archives

Can we connect local archives to institutions?

Can we use the political power of archives?

How do we make a community archive self sustainable?

How do we make archive a tool for healing and doing something useful?

How do we tag items in archives? For e.g,. do I allow for people to search for slurs or do I label it as violence? Do we censor the history or reopen up old history?

Do we translate our archives?

Localization, Local Content, Accessibility

Does online but non-open-web communication count? For e.g., is it enough to have rich and diverse Facebook or YouTube in your local language?

Can we localize open source content management platforms?

How do we create multilingual content that is accessible for people with disabilities?

How can we index (as opposed to archive) local content in meaningful ways to accurately represent the size usage of language?

How do you get from 300 Wikipedias (and only 100 really active) to 3000?

Epistemics, Frames, Ethics, Practices

How can we translate from dominant languages into marginalized languages without imposing foreign knowledge?

How are we framing the idea of decolonization? Looking at intersectionality and embodied and feminist and queer terminology used in the room; is the way we frame the problem itself problematic?

Can we problematize our preoccupation with measurement and content as a means of pursuing justice for indigenous languages?

Can we problematize the way we are speaking about language as content?

Consent, Licensing

How do we protect intellectual property, for e.g. community ownership?

How do we integrate indigenous feminist ethics around consent into how we think about consent in digital spaces?

Living in a colonial context with certain legal frameworks, what are the licensing schemes that balance privacy and openness and avoid extraction?

Finally, based on the questions and topics above, we moved on to identify some potential actions to decolonize the internet's languages, as well as the processes and resources that would be useful to achieve these actions.

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

Actions:

Build up more feminist discourse in African languages by:

→ translating the African feminist charter into more languages, enabling online action and dissemination.

→ producing 300 childrens books for early ages in 30 different marginalized languages, using 1.5k words that form the building blocks of feminist language.

Processes:

→ Get the right team of people together, needs to be able to deal with the activist and political space.

→ Train on context.

→ Build the terms, get good glossaries.

→ Find ways to make these materials oral as well.

Resources needed:

→ Animators, tech experts, graphic designers, project managers.

→ Consider publishing under CC license.

→ Funding for labor, meetups and dissemination.

Create training materials in Chatino language, an indigenous language of Mexico, so that more Chatino speakers can begin to access and create knowledge online.

Processes:

→ Decide on a group to start creating materials for what we know, for example teachers.

→ Respect orality.

→ Create materials that explain step-by-step how the platforms work.

→ Consult native speakers and share the information with other speakers, to engage community. Small groups can get a lot done.

Resources needed:

→ Use the software of open or free platforms, tuxpain, tuxmap, firefox, wikipedia, kiwix.

→ Review existing material with translation into indigenous language.

Center orality and oral languages when rethinking and designing online spaces, processes and experiences:

→ Have a low barrier of entry to increase participation. Spaces, such as archives, need to be motivated by people's daily needs, their concerns, motivations, desires and curiosity.

→ Create spaces that are heterogeneous and/or pluriversal. These spaces need to allow multiple processes of discovery and entry points.

→ Develop new ways to increase the discovery of oral knowledge that has been already documented and preserved in oral archives. For example, usage of audio-tags, emoji and cartoons that allow people to have a non-textual browsing experience. There is no need to type words or use a script.

→ Some of these actions also solves current accessibility and disability issues.

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

Processes:

→ Pilot a way to discover online content in an oral language like Zapotec.

→ Look at research on visual search methods, for example, consider how those methods could potentially inform oral search methods.

→ Oral language community to identify what they'd most like to make accessible online for a pilot.

Resources needed:

→ Money.

→ Researchers/techies.

Look to community led, digital archives to give us new and healing ways of understanding language. We want to:

→ Make community archives tools of transformation, supporting a sense of identity building and belonging, use as bridges within our communities and to other communities.

→ Document the process of archive-building, the dialogues we have about language, naming, etc.

Processes:

→ Reclaim public space and bring language and archives into the public space - looking at different examples: UK, Bosnia, Kashmir, etc.

→ Bring archives to different community spaces, including family spaces as much as organizational spaces, for discussions and exhibitions.

→ Learning from educational resources for the movement. Specifically looking at using different tech to have archives to speak to each other, open APIs, so more people can use them.

→ Look at the privacy side, putting a wall between consent and identity vs the artifact. Adjust the consent form so that things can be published after people's lifetime or consent can be withdrawn.

Resources needed:

→ Community-owned servers.

→ Internet access.

→ Mobilize diaspora for translation.

→ Identify gaps in content/software and better localise them.

Create new frameworks, practices and policy to negotiate consent, digitization control, privacy, permission and content removal from online platforms that's in-line with indigenous rights, knowledge and epistemic frames. Develop research protocols that provide a standard for an ethical baseline and more robust practices that's in-line with indigenous rights.

Processes:

→ Develop minimum baseline standards for ethical engagement with each other that uphold fundamental human rights.

→ Apply an intersectional lens: center input from indigenous and marginalized people discussing digitalization; plurality of process is central.

→ Look at existing tools, principles from marginalized and indigenous communities.

Resources needed:

→ Financial support, for e.g. fellowships to support people's time in marginalized communities.

Image by Aadele, CC BY-SA 4.0, via Wikimedia Commons

What's next?

This covening is just the beginning of our shared work to decolonize the internet's languages!

What participants are doing next:

→ Start the process to translate the African Feminist Charter and produce more children's books in African languages. Start with the concept notes we have created, move towards implementation.

→ Gather information about different experiences of indigenous speakers, including about challenges they face online with their languages. Possibly focus on children's educational material, Chatino community pilot.

→ Write a grant application for new modes of discovery of oral languages content with a Zapotec pilot.

→ Reach out to community-led archives and ask them to share their consent practices and their best practices for digitization.

→ Each create a list of 5 practices, 5 people, and 5 projects/ funders

- five examples from your work of best practices,
- five other people who can replace you, e.g. five other people who need resources, who may not be able to travel, but who need to be given spaces like this, and
- five fellowships to apply for or five projects that need funding.

Image by Aadele, CC BY-SA 4.0, via Wikimedia Commons

What Whose Knowledge? is doing next:

Much of what we've learned from participants at this convening will be rolled into our upcoming State of the Internet's Languages report, to be released mid-2020. The report will include quantitative and qualitative data (numbers+embodied experiences!) to demonstrate today's challenges and opportunities for diversifying languages online. We see this being used as a baseline to build awareness and further agendas for action with marginalized communities, open knowledge folks, and tech communities to decolonize the internet's languages.

What YOU can do next:

We invite techies, researchers, and communities to also wrestle with these challenges, opportunities, and questions to decolonize the internet's languages. Please join in and help carry conversations and actions forward!

Dreams for the future:

Participants at this convening said they would love to have an ongoing multilingual space to discuss Decolonizing the Internet online, where their communities and others can join in. Perhaps a wiki? Whose Knowledge? sadly doesn't have capacity or resources to make this happen right now, but we would love to someday have this dream come true. Let us know if you want to help!

Join the conversation!

You can use

#DTILanguages
#DecolonizeTheInternet
#DecolonizingTheInternet

to tag your conversations, questions, ideas and suggestions on social media platforms.

Sign up for the Whose Knowledge? Newsletter to receive monthly updates and suggestions for how to get involved.

Reflections

Here is what some participants have said about this gathering:

Wangui Wagoro

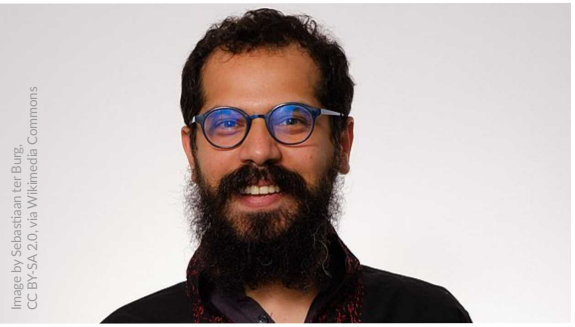
Image by AWDB-wikiwomen, CC BY-SA 4.0, via Wikimedia Commons

To discover there is a whole community of people who are ethical, equally concerned about the same things I have been trying to grapple with for a long time alone and we can just speak the same language, without justifying ourselves or explaining what the concerns are (...) so for me it's very rewarding to see that I am not alone, I am not crazy and there is a big issue that we are at the pinnacle of, pioneering the revolution.

I would like to see the community of practice grow. However we do it, whether consciously or unconsciously that we take these messages, that we keep in touch and that we find physical spaces to share what we've shared here. I would like to see the community of practice grow and I want to bepart of making it grow and I want to be part of it going forward.

Ana Alonso

I think that this is a great opportunity to hear what other people are doing in other parts of the world, and my main interest in coming to Decolonizing the Internet's Languages is to really think about what needs to be done, or how we can direct the internet we have today and make it accessible to indigenous languages and indigenous people that are not represented on the internet. I'm trying to think about how we can improve the internet and how we can make it accessible for non written languages.



Majd Al-Shihabi

The experience has been very inspiring because I know that there are a lot of people working and struggling with the same questions, but at the same time knowing that intellectually, and feeling that, and seeing that in person is a completely different experience, and is very reaffirming that yes I'm asking the right questions, yes there are people who are in solidarity with me, who will help me out in my process of discovering my way or my communities' way of doing decolonization, so meeting these people in person has been invaluable.

Subhashish Panigrahi

This conference is sort of a bridge to bring people together and enable them to work together, because the community that really is contributing towards development of languages and ensuring that communities make use of their own languages on the internet (...) it's really a small community, so I think bringing so many people together just sorts of opens that door to work together, and also share the issues that they are experiencing in their own work and sometimes that is very important.

Solana Larsen

Part of the challenge is to keep explaining why it's so urgent for these challenges around language, utility, and co-creation of the internet to be overcome. In this group, we had a really strong shared understanding of why language matters and how it relates to de/colonization and why it even makes sense to speak of such terms in relation to the internet, but so many people (affected or not) still lack the inspiration to think beyond social media when it comes to use of the internet and perhaps also the skills to reimagine better online communication for the future.

Image by Aadele, CC BY-SA 4.0, via Wikimedia Commons

Francesca Beard

To be part of this gathering of people, intent on imagining a real difference, on changing the course of our present story towards a kinder, more positive history felt so vital, at times, the energy in the room had an electric presence. Whose Knowledge? and Decolonizing the Internet's Languages are doing vital, strategic, necessary work for humans everywhere. They are mapping out not only the deep and real problems but also the deep and real opportunities of our digital revolution, leading a call to arms to fight for an internet which supports human connection and allows us, as individuals, to access a wider understanding of who we are, by listening to different stories and the knowledge that is found in the spaces between the words.

Image by Heykc, CC BY-SA 4.0, via Wikimedia Commons

Amira Dhalla

The Internet is such a fundamental part of our lives, and yet the majority of it is written in English. This is not reflective of a diverse world and inclusive of the many communities or areas or languages that are present. This Decolonizing the Internet's Languages brought together a diverse group who was able to discuss the different issues and inspire each other through each individual collective action to decolonize languages online. Being in the room was heartwarming, supportive, and encouraging. For me, it confirmed why we need events like this and the importance of future ones.

Jessica Horn

Image by Wotancito, CC BY-SA 4.0, via Wikimedia Commons

I think the most valuable aspect of this space is the fact that it's interdisciplinary and very diverse in terms of where people are coming from with their own knowledge bases, their own political analysis, situations, geographically, linguistically, culturally, etc., and so for me that diversity is beautiful,hat's also interesting is that already as we are identifying the kind of visions of the ethical basis upon which we think a decolonized internet would exist. People who've never met each other were coming up with a lot of common points, which means that actually we share a broad sense of direction in terms of where we think this should go. So for me that makes me quite hopeful because it means that the constituencies that we come from are already thinking about what we want, and we're communicating this as we're sitting here together. So I feel like there's actually a constituency already pushing for this notion of a decolonized internet, we're already thinking about it. Because we already have the pace, we're already moving. I feel there's a possibility that it can happen.

Gratitude

Decolonizing the Internet's Languages would not have been possible without everyone who joined the conversation. Apart from being forever grateful to our incredible participants, we want to take this opportunity to thank a special few of our supporters, partners, friends and allies who made the conversation happen, and who will keep the conversation going.

Facilitation
Francesca Beard

Logistics
Irene Tait

Design Co-conspirators
Amira Dhalla
Kelly Foster
Mandana Seyfeddinipur

Whose Voices? podcast support
Jake Orlowitz

Mozilla Foundation
Sarah Allen
Lucie Click
And everyone on the MozFest and
RSA House teams for being great hosts!

The Citadines Hotel, Holborn
The entire team at Citadines for accommodating us so comfortably and warmly!

Funders
Shuttleworth Foundation
Mozilla Foundation
Open Society Foundations

Whose Knowledge? works with marginalized communities around the world - including women, people of color, indigenous peoples, LGBTQIA communities and folks from the Global South - to build an internet for and from us all. We may be marginalized by structures of power and privilege, but we are the majority of the world. Our knowledge is the knowledge of most of the world.