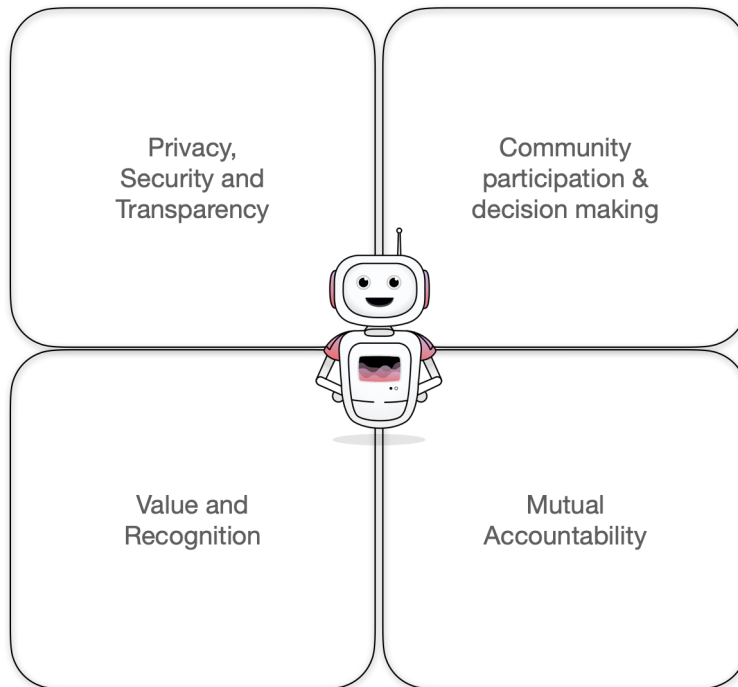


# Mozilla Common Voice Governance Doc V1.0

## Common Voice

**moz://a**



### Privacy, Security and Transparency:

We are committed to the privacy and security of everyone who takes part in Common Voice. Our policies and practices respect all people's data privacy rights and we try to go further than the minimum legal requirements to earn your trust. We **pseudonymise** published voice data, provide clear pathways to have your profile deleted and data **deleted** from future releases, and collect only enough information via **analytics** to ensure strong platform performance. If you have **do-not-track** set up through your browser, you'll be excluded from our site analytics. We do **not use third party cookies**. If we ever change anything in our policies, we will **let you know**. If you want to give us **more information** about

yourself - such as your email or your age - that's really helpful for understanding and improving our dataset. However, it is **completely optional**. You don't need to create an account to participate. See our privacy policy and terms and conditions for more information.

In 2021 we established our first **Diversity Datasheet framework and goals**, which looks at demographic representation in MCV. We will publicly release a report on our progress in 2022.

Have a question about any of this? We'd love to hear it on [commonvoice@mozilla.com](mailto:commonvoice@mozilla.com)

## Community participation and decision making:

The direction of the platform, including guidelines, policies and practices, is established by the **Common Voice Product Team at the Mozilla Foundation**, and informed by their ongoing engagement with the Common Voice community - the people who give up their time to make Common Voice possible, whether by contributing their voice, sentences, or code, or by using the dataset to build models or tools. A **Language Reps Council** represents the language communities who make Common Voice possible, and they meet at least monthly to engage on current issues, and bring discussions to their wider communities. **Dataset consumers and funding partners** also meet at least quarterly. The input of these groups, as well as casual contributors who share feedback in our forums, is **treated equally** in prioritization, which is guided by our [values-driven matrix](#). When we establish a guideline - for example defining language for the purposes of the dataset - we will always make space for **open conversation** about exceptions, for example to take account of political and social history.

There are also **multiple asynchronous pathways** to suggest changes, flag issues and share ideas (public roadmap, github backlog, discourse, matrix, AMAs, open events), which are fully considered by the team. For our regional and language

specific work, for example in East Africa, we also work with **in-country community champions** to understand their unique needs and concerns.

In 2022 we are launching an **Advisory Panel**, which will bring together experts in social justice, including in issues of race and gender, computational linguistics, AI ethics, and endangered languages to support some of our most important decision making.

## Accountability:

Accountability goes in both directions. We are truly open source - we welcome all languages, the platform code is available under an MPL2.0 license and the dataset is available under a CC-0 license. We have reasonably few rules that require **enforcing**. To the extent they do, they are rooted in protecting people from harm (see: community participation guidelines, age guidelines, unacceptable usage). If we identify an issue - either directly or via a community member bringing it to our attention - we will **first** flag it and **educate** the person/people on the importance of the guideline, the **second** step is to give a written **warning** that explains what will happen if the behavior persists, the final is to **remove** their accounts from the platform and its associated spaces. \*Note this is the pathway for governance issues, and does not usurp existing community escalation pathways.

If a community or member feels we are not respecting our data governance principles, they can at first **submit the concern** in writing to the **Mozilla Foundation Product Team**, who will give their complaint full consideration - and where appropriate, discuss with the **Language Reps** council - and deliver a written response. If the community is not satisfied with the response, and would like to escalate it further, the responsible **Vice President** at the foundation will be asked to review. The final escalation path is the **Steering Committee**. This is composed of executives from Mozilla Foundation and Common Voice advisory and funding partners, and/or Mozilla's **Legal** team if appropriate; they will review their concerns, and respond as they deem fit. Where possible, we will be transparent about the process so that others may learn - without, of course, violating people's privacy.

## Value and recognition:

Common Voice is an investment in the future. **Technology that is more equitable and inclusive is a service to all of humanity.** When people contribute their voice, they are part of that better future. As more sectors leverage voice technology to provide services - from banking to medicine to government - this equity becomes ever more crucial.

We know that different communities have very **different contexts** - some communities have already had so much taken from them, and received only exploitation and oppression. In 2022 we are working with indigenous and endangered language communities, as well as social justice experts, to create participation pathways that **give them more control** over their data and how it's used. To get involved, please reach out to us on [commonvoice@mozilla.com](mailto:commonvoice@mozilla.com).

We know that equitable AI can seem a long way off, and we want our community to get value from being part of Common Voice ahead of that. We have a **reward and recognition** scheme that helps provide a platform for contributors (Mozilla blog and twitter takeovers), provide them with skills (free training and learning) and makes participating fun (laptop stickers, community socials, shout outs and swag!)

We know that time-poverty is a real barrier in open source, and that it particularly affects those with disabilities, economic difficulty and care responsibilities, so we are also setting up an **enablement** track that provides things like computer processing power and mobile internet data to people who would like to help but cannot give up their time for free.

In particular in our underrepresented/underserved languages work, we take on **partnership models** with knowledge holders such as tertiary institutions, linguist communities and media houses. We **publicly recognise** their engagement - e.g. sentence contribution - as partners on the work, set up **skills initiatives**, provide funding and support **connecting activities** to improve their access to other like-minded holders of local knowledge.

We also **share information**, so you can gain insights from our work - we publish as much as we can about our methodologies, results and findings, as well as making some data analytics public. If you'd like to collaborate on research, reach out here [commonvoice@mozilla.com](mailto:commonvoice@mozilla.com).

## **Learn more**

- [CC0 licenses](#)
- [Privacy Policy](#)
- [GDPR](#)
- [Language definitions](#)
- [Language Reps](#)
- [Reward and recognition](#)
- [Community participation guidelines](#)