

An account of how critical information about privacy is missing in indigenous languages of India and ways to avoid it

Subhashish Panigrahi

Abstract

As of the September 2020 statistics, 93% of the 1.3 billion population of India has enrolled for Aadhaar, a national digital ID. However, the website and other resources maintained by the government are only available in 12 Indian languages which does not include any of the indigenous languages. There are more than 200 indigenous languages that are spoken by over 106 million people (8.6% population). While collecting subjective narratives for a documentary film, I came across an emerging trend that cuts across the indigenous and other minority language speakers: most people who enroll for Aadhaar are not provided with any critical information pertaining to their rights in their native languages. Most individuals enroll as linking of Aadhaar with public benefits (like ration or pension or even banking) is enforced by authorities and biometrics captured through Aadhaar are used for authentication. I will indicate how the lack of multilingual and universally accessible resources lead to ethical and moral exploitation of indigenous peoples. I will cite case studies of adult individuals who have consented to appear in the documentary to elaborate the dearth of specific kinds of linguistic and informational resources, and will indicate the kind of such resources as a recommended list for similar public services.

In the last decade, many governments and public administrations in countries with poor freedom indices have started to roll out programs that collect and store personal data. Lack of legal and regulatory frameworks endanger the freedom of many indigenous peoples who are already vulnerable. Lack of creation and dissemination of critical public information in different spoken languages and in accessible formats furthers such vulnerabilities.

This paper will assess the needs for different kinds of resources that the anecdotal findings indicate. It will highlight some of the interviewed representatives that represent indigenous communities and other minority language speakers. Current and potential practitioners of indigenous and minority/marginalized native languages who work in complex geopolitical and socioeconomic environments, and emergency response like COVID-19 outbreak will be benefited from this work. Those who work at the intersections of languages, social justice and rights will also be able to make

use of the learning and can contribute for further collaboration on building resources and frameworks for resource creation.

Non-verbatim transcription of presentation

Hi, my name is Subhashish. I was a digital identity fellow at Yoti during 2019 and 2020, and I researched on the exclusion of some of the marginalized communities because of Aadhaar—India's biometric ID. Disseminating public information in multiple languages has always been a challenge especially in a country like India that has a population of 1.3 billion people that speak more than 780 languages. Out of that, there are 200 plus indigenous languages known as Adivasi languages that are spoken by multiple ethnic groups—about 106 million people in total that constitute about 8.6 of the population. That's slightly outdated data. In 2021, we'll have a census that will provide a little bit more insight about the linguistic diversity and the number of Adivasi communities. In 2009, the Indian government launched the Aadhaar program to collect a lot of personal data including biometrics and that includes fingerprints, iris, photograph, and, all of that data were stored in a centralized database. The biometrics authentication is used used for public welfare including food, ration, and now there is a discussion to provide vaccination using the Aadhaar as an authentication method.

The question here is why one should worry about the lack of literacy about a particular ID program and how that could lead to exclusion. That was one of the questions that I included in my Fellowship research. The way I documented some of the exclusions was through creating a documentary film called “MarginalizedAadhaar”. So, that was a documentary research looking at the subjective narratives. The people that I interviewed mostly were indigenous language speakers or speakers that speak different oral languages, dialects or predominantly oral languages. I interviewed people that speak such languages and have disability. That is an additional level of challenge with access. There are many people that I interviewed are monolingual.

[Ramani speaking Sora]

How do we do it? what do we do [enroll]? They call this "Aadhaar card". How do we do it [enroll]? We need to get our pictures clicked. They told that we need to get our pictures taken. What will they do with this Aadhaar card? They are saying that one can get anything [with Aadhaar card]?

[Subhashish]

Many people also are illiterate in official languages. To provide a little bit of context, 93% of India's population has enrolled Aadhaar. And Aadhaar website—the UIDAI website that hosts all the information about Aadhaar and other necessary information about Aadhaar—is now available in 12 languages whereas most of the federal government sites are available only in two languages. And

that's that's really an inclusive way to begin with. However, not a single of those 12 languages include an indigenous language. There have been cases of denial for food or ration or relief packages during the onset of the pandemic where biometric id were were used for authentication. So, Aadhaar was used for authentication in this case and biometric means were used for this authentication process. That definitely posed more challenges. The problem with biometrics is it is a very consentless and oppressive technology. I Professor Reetika Khera—who is an economist—and Sunil Abraham—who is a technology researcher—have shared their share of concerns about this challenge.

[Reetika Khera]

What the government did is—program after program, scheme after scheme—they started saying that not only do you have to have Aadhaar, but you should also link up your Aadhaar with the registry of that program with the database that exists with the government for that program. So, that again has become a huge source of exclusion because sometimes the linking doesn't happen successfully. And in many programs the linking is actually a two-step thing—one is to link it with the computer database of that program. But, also, if it's a cash based transfer or cash support, then they have to also link their Aadhaar number with their bank account. That has also created a big mess in the banking system because the banking system is itself mapping on to the Aadhaar payment bridge system.

[Sunil Abraham]

The problem with biometrics is—by definition—it is consentless technology. It doesn't need your consent for it to work, especially fingerprints and to some degree facial recognition, and iris matching as well. So, it violates the most important principle in privacy law, and in the broader movement of privacy protection—which is protecting consent. The second is, it is really an unempowering technology for the people that use it.

[Subhashish]

Now, there is a lack of active consent seeking when it comes to biometrics. People that are illiterate or monolingual are asked to comply to enroll or to authenticate. The the bigger question here is—why a government program collects so much private data, especially biometrics, when the purpose is not border control or security, but delivery of food grains or ration. So, if we take a step back and talk about governance—governance is not transactional. Governance is merely the process of upholding the rights by the state for the citizens. Privacy is a fundamental right. So, when there is a public program that violates privacy, knowingly or unknowingly, there is a case of exclusion. One

of the emerging trends that I saw is there is no literacy program for digital rights. Digital rights are understood as the human rights in relation to the internet and other digital interactions. There is no focus also on literacy of constitutional rights. In most cases it's compliance. So, a researcher who has done a lot of ground research tells me that the role of Aadhaar for the beneficiaries—the way they perceive it—is to largely something that is just to comply with the government requirements. So, we see a lot of forced use of biometrics with no digital literacy. Prathamī—another a community member from the Naik tribe in Goa in western India—tells me that elders did not get any access to the internet.

[Prathamī]

My generation had a lot of advantages. Those before me, one decade before me—the elders—did not get any access to the internet.

[Subhashish]

Many elders had to enroll for Aadhaar. Illiteracy in official languages, and illiteracy in general is a huge challenge particularly in indigenous groups. The literacy rate among Adivasis, mostly indigenous communities, is about 40 percent. It's much more less for females—about 37 per cent. So, I interviewed several community members including a community elder from the Lanjia Sora community. Lanjia Sora is a dialect that's predominantly oral. This community elder—who is illiterate and is monolingual—tells me that they were asked to enroll for their Aadhaar. The government asked them to enroll. She was never told the reason why they were asked to do so. So, she never understood the real reason for enrolling for Aadhaar. Then there are monolingual indigenous speakers. And when the goal is to collect massive amount of data and create a huge database, the collection process is never inclusive and it's never meant to educate people but to enforce or sell [an idea]. Usha Ramanathan, a law researcher details a lot about that. Professor Mandana Seyfeddinipur, a noted linguist who I interviewed, underlines the linguistic role—the need for linguistic diversity and creating resources, and how that impacts the critical public information dissemination.

[Prof. Seyfeddinipur]

If we are in a situation where the government information, the resources that are available to people in terms of education, health, access to resources, is in a particular language that they cannot access, then that means that it's a measure of exclusion.

[Subhashish]

I've seen that exclusion in many levels. One particular incident comes to my mind—I was in a remote village in the eastern part of India in a state called Odisha. A community elder there tells me that they have a system where someone called Endia, a misnomer for India. This person [endia] is modern day messenger on behalf of the government who comes and announces in the village on a regular basis orally about government program deadlines and the community members have to comply with these announced deadlines. This process is not to educate people but to tell them that they have to go and enroll or they have to go and change their details if there is any error and so on. The local authorities always made it look like Aadhaar is mandatory for public welfare. This is a reason where there is a lot of exclusion that we have seen. So, the forced consent—that many illiterate people, disabled people and other marginalized communities have gone through—has led to a lot of exclusion in India. And I would actually argue that this exclusion is much more for people that are indigenous and speak different indigenous languages as indigenous communities have a higher illiteracy rate and they are monolingual. If they have a disability, that's an additional level of exclusion. When the community members that are illiterate or monolingual or disabled, they rely a lot on family members. The family members may be literate and multilingual but there is hardly much educational content that is created to educate people. So, even though a family member is literate, they might not have much idea about how their privacy is in question when they enroll for Aadhaar or when they use Aadhaar for authentication. So, people rely on public authorities. As I said already, it is based on compliance and not education.

What could have been a better approach? And many interviewees have shared their share of suggestions and recommendations. The first thing is literacy of rights— be it human rights or digital rights—particularly privacy. People need to be educated about those rights, and that has to be done before massive enrollment. Secondly, there has to be a survey to learn how many people are monolingual and how many people are multilingual. If there is a need for certain language resources to be created, to provide literacy for people, that has to be done first. These government programs, particularly the ones like Aadhaar, have to have a lot of Open Educational Resources (OERs) that are basically learning resources with open licenses. These OERs need to be disseminated publicly and they have to be in accessible mediums. Some examples could be creating audiovisual resources so people who are illiterate or visually impaired or are oral language speakers can access such resources. The resources also need to be captioned so that people with hearing impairment or people who are multilingual can watch them. The fourth recommendation is stopping or regulating the use of biometric IDs—how it should be used if it should be used. The first thing is—it has to be limited or stopped because it is an oppressive technology. It is a faulty technology in a way. The biometric ID should not be used in a mass scale, particularly in rural communities that are already marginalized in many ways, and particularly for social benefits when authentication is made using Aadhaar. We have seen that the people who are below average poverty line are mostly marginalized, and many of them are illiterate. It increases their marginalization. So, to provide a little bit of insight on that, a 2018 the State of Aadhaar report, a research that was done by IDInsight, shows that only 0.8%, 2.2% and 0.8 % respectively of the rural Public Distribution System (PDS) beneficiaries respectively from the Indian states of Andhra Pradesh, Rajasthan and West Bengal were confirmed

to be excluded from their entitled benefits. Aadhaar was the primary issue for those exclusions. The same report also tells that seven percent are only aware of the entire Aadhaar authentication process that is linked to their privacy. Smart cards, on the other hand, could be a better alternative to biometric IDs and they could use modern encryption. They could use infrastructures like the public key encryption that provides a lot of sovereignty to the users. It also involves a layer of informed consent. So, the users decide whether or not they want to enroll or whether or not they want to provide authentication. These recommendations might sound very specific to India but they can be relevant to elsewhere, particularly when it comes to indigenous groups and their rights. It is very important that their privacy and their rights are kept in the focus, and they are given the utmost importance. Thank you so much for listening to me!