



CALL FOR INPUTS BY THE UNITED NATIONS

Study On Indigenous Peoples Right To
Data, Including Data Collection And
Disaggregation

January, 2025

Call For Inputs By The United Nations

Study On Indigenous Peoples Right To
Data, Including Data Collection And
Disaggregation

Authors: Aryan Rana, Chirkankshit Bihari Bulani, Aadit Seth, Arshmeet Singh, Yug
Raman Srivastava

Research Consultants:

Prof. (Dr.) Jai Shankar Singh, Prof. (Dr.) Yogesh Pratap Singh, Prof. (Dr.) Naresh Vats,
Dr. Ivneet Walia, Mr. Abhijeet Srivastava



CENTRE FOR ADVANCED STUDIES IN CYBER LAW AND ARTIFICIAL INTELLIGENCE [CASCA] is a research-driven centre at RGNUL dedicated to advancing scholarly research and discourse in the field of Technology Law and Regulation. As a research centre of a leading institution in India, we are committed to promoting interdisciplinary research, fostering collaboration, and driving innovation in the fields of cyber law, artificial intelligence, and other allied areas.

For more information

Visit cascargnul.com

Disclaimer

The facts and information in this report may be reproduced only after giving due attribution to CASCA.

Introduction:

This submission by **Centre for Advanced Studies in Cyber Laws and Artificial Intelligence (CASCA)** at Rajiv Gandhi National University of Law, Patiala, India is in response to **Call for Inputs: Study on “Indigenous Peoples right to data, including data collection and disaggregation”** by Expert Mechanism on the Rights of Indigenous Peoples. **Centre for Advanced Studies in Cyber Laws and Artificial Intelligence** is research centre established under Rajiv Gandhi National University of Law, Patiala and fosters research in the intersection of cyber laws, artificial intelligence and society. This submission is part of CASCA’s initiative is to integrate Artificial intelligence with public policy to contribute to a more inclusive and just society.

This submission shall serve to establish the following:

- Further the case for integration of artificial intelligence in traditional methods of data collection used to collect indigenous and tribal data.
- Provide insights into AI usage and its integration through examples of AI usage in data collection of tribes of India.
- Establish that such data collection must accommodate Free, prior and informed consent and respect data sovereignty of tribals individuals.
- Elucidate upon challenges and solutions in integration of AI in indigenous data collection

Objectives and Challenges of data collection for indigenous people

The considerable objectives for data collection concerning the indigenous people i.e. identification of socio-economic disparities such as income, education, and health, along with a way for these disparities to create a basis for targeted intervention. The Indigenous and Tribal Peoples Convention, 1989 (No. 169)¹, whose Article 24² and 25³, which provide for social security and health, and mandate that social security and health services be extended to people of tribal origin without any discrimination. Article 33⁴ mandates the government entities or agencies for effective administration and realisation of rights granted to tribals under the convention. In India, the tribal and indigenous population has received recognition by name of ‘Scheduled tribes’ under Article 342 of the Indian Constitution⁵. Under Article 15 of the Indian Constitution⁶, the State of India is empowered to undertake affirmative action regarding groups falling under the categorisation. Further, Right to privacy, which is protected internationally under Article 12 of UDHR⁷ and Article 17 of ICCPR⁸ and domestically in India under the Article 21 of

¹ International Labour Organization, Indigenous and Tribal Peoples Convention (No. 169), adopted June 27, 1989, entered into force Sept. 5, 1991, 28 I.L.M. 1382 (1989).

² Indigenous and Tribal Peoples Convention (No. 169), art. 24, June 27, 1989, 28 I.L.M. 1382 (1989).

³ Indigenous and Tribal Peoples Convention (No. 169), art. 25, June 27, 1989, 28 I.L.M. 1382 (1989).

⁴ Indigenous and Tribal Peoples Convention (No. 169), art. 33, June 27, 1989, 28 I.L.M. 1382 (1989).

⁵ Constitution of India, art. 342.

⁶ Constitution of India, art. 15.

⁷ Universal Declaration of Human Rights, art. 12, G.A. Res. 217A (III), U.N. Doc. A/810 at 71 (1948)

⁸ International Covenant on Civil and Political Rights, art. 17, Dec. 16, 1966, 999 U.N.T.S. 171.

the Indian Constitution⁹ envisages data sovereignty for these marginalised groups. Given that right to data sovereignty is protected, there must be the involvement of data in the policy formation along with the allocation of resources to create a basis for an informed decision with evidence-based insights, to promote equitable growth for indigenous people to solve their unique challenges. Thereby, it is emphasized that how the data should be collected must preserve and promote indigenous culture and traditions by documenting languages, and practices, to ensure that they are properly protected and valued within their different contexts,¹⁰ hence the indigenous priorities and needs are culturally preserved and also inclusivity, diversity, and collaboration needed for the proper reflection.

AI Integration as a solution to these Challenges

The challenges of preservation of indigenous cultures, their languages and way of life can be appropriately addressed by means of Artificial intelligence. Artificial Intelligence possesses the capability in transcription and translation and can be exhaustively used in data collection on indigenous and tribal groups.¹¹ Other challenges such as cost-effectiveness, accessibility can be countered by deployment of algorithms and their use in data collection.¹² For elucidation with an example, The Toda Ashram School situated in Etapalli, Maharashtra in India has used an AI software to combat Malnutrition among the tribal student population of Gadchiroli.¹³ The software captured pictures of students with their plates and analysed the nutritional content of the food so as to figure out the lacunae in the diets of students. Such innovative methods of data collection can address cost effectiveness and accessibility at the same time.

AI Integration with Data collection : A Case Studies from India

Case Study 1: Platform by ICMR for Surveillance of food and water-borne diseases in North-East India

Studies show that globally, nearly 600 million people are affected due to food and waterborne diseases causing more than 400,000 deaths every year.¹⁴ In India, studies show that due to lack of data, its nearly impossible to correctly estimate how widespread the problem is.¹⁵ In this regard, the Indian Council for

⁹ Constitution of India, art. 15. And Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 S.C.C. 1 (India).

¹⁰ Leanne R. Simpson. "Anticolonial Strategies for the Recovery and Maintenance of Indigenous Knowledge." *American Indian Quarterly*, vol. 28, no. 3/4, 2004, pp. 373–84. *JSTOR*, <http://www.jstor.org/stable/4138923>. Accessed 21 Nov. 2024.

¹¹ Erin Kalejs, *How AI is Helping Indigenous Languages Survive and Thrive*, AI FOR GOOD BLOG (Aug. 9, 2022), <https://aiforgood.itu.int/how-ai-is-helping-indigenous-languages-survive-and-thrive/>

¹²World Bank, Use of AI Technology to Support Data Collection (2021), available at https://gps.worldbank.org/sites/gps/files/knowledge_products/2021/Use%20of%20AI%20technology%20to%20support%20data%20collection.pdf.

¹³ The Better India, IAS Officer Uses AI to Fight Malnutrition in Maharashtra's Govt Schools, available at <https://thebetterindia.com/338306/ias-officer-shubham-gupta-used-artificial-intelligence-to-fight-malnutrition-govt-schools-maharashtra/>.

¹⁴ 3. World Health Organization (2022). Food safety. Available at: <https://www.who.int/news-room/fact-sheets/detail/food-safety>. (Accessed July 26, 2024).

¹⁵ David S. Wilkins et al., The Role of Community-Based Interventions in Reducing Hypertension in Rural India, 115(10) *PMC 1007* (2005), available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC1151007/>.

Medical research initiated a digital data platform combined with advanced analytics to predict the outbreak of the disease.¹⁶

The study developed a digital platform to monitor food- and water-borne pathogens in Northeast India, utilizing advanced technologies to enhance public health. Key steps included firstly, data collection, involving Real-time sampling of food, water, and clinical specimens from regional markets, water supplies, and healthcare facilities. Secondly, Data Integration was done using IoT devices, remote sensing, and mobile-based applications. Thirdly, Artificial Intelligence was employed and Algorithms analyzed pathogen trends, predicted outbreaks, and identified correlations between environmental factors and disease. Additionally, Geographic Information System (GIS) mapping visualized pathogen hotspots for targeted intervention. Predictive analysis was done to recognise zones of risk on basis of climate and socio-economic factors. Lastly, Insights from analytics informed regional health strategies and stakeholder actions.

Case Study 2: Predicting child anaemia in the North-Eastern states of India: a machine learning approach

Anaemia is a condition in which the number of red blood cells or the haemoglobin concentration within them is lower than normal. It mainly affects women and children. Globally, it is estimated that 40% of all children aged 6–59 months, 37% of pregnant women and 30% of women 15–49 years of age are affected by anaemia.¹⁷ The prevalence of anaemia in India among six groups as per the National Family Health Survey 5 (2019-21), is 25.0 percent in men (15-49 years) and 57.0 percent in women (15-49 years). 31.1 percent in adolescent boys (15-19 yrs), 59.1 percent in adolescent girls, 52.2 percent in pregnant women (15-49 years) and 67.1 percent in children (6-59 months).¹⁸ A study published in International Journal of System Assurance Engineering and Management, which is the official publication of The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden.¹⁹ The study, conducted by researchers in Northeast India, focused on anemia among 8,514 children aged 6–59 months using data from the National Family Health Survey-4 (2015-16).

Researchers used machine learning techniques, including ridge regression, LASSO, and elastic net, to identify key predictors of anemia, such as maternal anemia, child's age, maternal education, and wealth index. These ML models achieved over 70% ROC values, showcasing their ability to handle complex

¹⁶ Ashok Kumar et al., Developing a Digital Data Platform for Surveillance of Food and Water-Borne Pathogens in North East India: Insight for Public Health Advocacy, FRONTIERS IN PUBLIC HEALTH (2024), available at <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2024.1422373/full>.

¹⁷ World Health Organization, Anaemia, May 1, 2023, available at <https://www.who.int/news-room/fact-sheets/detail/anaemia>.

¹⁸ Press Information Bureau, Ministry of Health and Family Welfare, Anaemia Mukht Bharat (Feb. 4, 2022), available at <https://pib.gov.in/PressReleasePage.aspx?PRID=1795421>.

¹⁹ Meitei, A.J., Saini, A., Mohapatra, B.B. *et al.* Predicting child anaemia in the North-Eastern states of India: a machine learning approach. *Int J Syst Assur Eng Manag* **13**, 2949–2962 (2022). <https://doi.org/10.1007/s13198-022-01765-4>

datasets and identify multifactorial relationships more effectively than traditional methods. The study's outcomes highlighted the value of AI in improving public health strategies by enabling targeted interventions and resource allocation. For instance, maternal anemia emerged as a critical factor, emphasizing the need for comprehensive maternal and child health programs.

By leveraging predictive capabilities, the above-mentioned studies have demonstrated how AI can enhance decision-making and policy formulation in public health. These capabilities can be further used to facilitate data collection of indigenous groups by improving access to technology.

Shortcoming, Challenges in AI Integration in Data collection on Indigenous People and Solutions

1. Technology Stigma and Apprehension

It is established that even though integration of artificial intelligence comes with significant advances, however such an integration may compromise on some foundations of sound data collection. With the rise of fears among general populace in regards to privacy breaches and unauthorised data collection²⁰, this trend could worsen when dealing with already stigmatized indigenous communities. It is also imperative to take note of the fact that the primary apprehensions to technologies based on data stem from training of such technologies on incomplete data sets, giving rise to historical inequalities. It may also lead to the increase in the data manipulation along with, it increases the risk of cyberattacks, broadening the digital divide, and may also lead to the identity theft and giving rise to the ethical concerns. Moreover, the integration of the modern innovation is also expensive during its initial phases of implementation and sometimes may not store large amount of data.²¹

2. Data Content limitations

Further, the data collected from the indigenous peoples the one that the developers of data collection deem important. For example, the Scheduled Tribes (ST) population data collected through the Census of India. While it provides demographic details such as population size, literacy rates, and employment statistics, it does not capture the impact of factors like traditional cultural practices, mental and emotional wellbeing, access to land, or the role of indigenous knowledge systems on health and community well-being. There also exist a chance of data cleansing which may lead to an incomplete and missing data, this could happen both in the preliminary stage of collection or after the advance stage of the data collection. Then, due to the complications in the modern innovation framework, it leading to the duplicate entries²² as well as doubting the accuracy and raising concerns relating to ethics and privacy.

²⁰ The Economic Times, 60 Online Users Fear Unauthorised Data Collection, Only 11 Users Read Privacy Policies: Survey, economictimes.indiatimes.com (last visited Jan. 26, 2025).

²¹ Institut Européen de Droit, *The Pros and Cons of Using the Latest Technology for Data Collection, Storage and Extraction*, ied.eu (last visited Jan. 26, 2025).

²² FasterCapital, Limitations of Data, fastercapital.com (last visited Jan. 26, 2025).

Solutions:

1. There should be an active engagement of Indigenous people at all stages for proper data collection which includes Free Prior Informed Consent (FPIC) so that the communities are informed about the elements of sound data collection which include purpose methods and potential uses. The related methodology must also address different challenges and respect the different cultural norms. It also includes multiple indicators which reflect the diversity of experiences within the different groups.
2. There must also be a sound and ethical data-handling process including key aspects of data sovereignty along with maintaining the required confidentiality and the findings in accessible formats. This process also includes multiple aspects such as reaching remote populations, and overcoming technological barriers, to encourage effective participation.
3. Ethical collection of data must also ensure FPIC. All the indigenous people should be informed on the purpose, procedure and the possible impacts of collecting data. This principle establishes trust, plus ensures that such initiatives are subjected to community priorities. Peoples have the right to approve or reject any data collection initiatives that concern them. In practice, FPIC is frequently ignored or inadequately implemented, particularly in research conducted by state entities or commercial entities. These protocols should outline how consent is obtained, the information provided to communities, and the mechanisms for challenging any data collection project that is conducted without proper consent. There should also be an enforcement in legal frameworks that protect Indigenous Peoples' rights to FPIC, ensuring that data collection initiatives cannot proceed without explicit consent from the affected communities.

Conclusion

This response paper, by means of case studies discussed and concepts established, tried to establish the shortcomings of data collection on indigenous people and established the potential advantages of AI integration in data collection practices established for such collection on indigenous people. AI can bridge community gaps and address challenges of accessibility, effectiveness and participation. Additionally, introduction of algorithms in data analysis of data so collected can attribute to effective and inclusive decision making and policy interventions. However, the shortcomings of AI which include unauthorised data collection, mismanagement and misuse must be dealt with accordingly while designing such data protocols.