

Title: Beyond Citizen participation and Statistical Integrity: Ethical Considerations in Harnessing Citizen-Generated Data for Official Statistics

Citizen-generated data (CGD) has become an increasingly important resource for enriching official statistics, particularly in contexts where traditional data collection methods face limitations of cost, timeliness, and coverage. In Kenya, the growing reliance on CGD reflects both the promise of participatory approaches and the challenges of embedding them within National Statistical System. This paper, interrogates the ethical foundations necessary to ensure that citizen contributions are not only statistically sound but also socially just and credible.

We argue that data ethics must extend beyond technical concerns of quality to encompass respect for citizens as subjects, not merely as data points. Central to this is the principle of informed consent: citizens must be fully aware of how their data is collected, processed, and used, with clear opportunities to opt in or out. Integrity in collection processes is equally vital, requiring transparent methodologies that avoid manipulation, coercion, or selective framing. Ethical practice demands that data handlers such as; civil society organizations, community groups, and other intermediaries, demonstrate credibility, accountability, and independence. Without these safeguards, CGD risks being skewed to advance narrow interests rather than reflecting collective realities.

The paper further emphasizes reciprocity: data must ultimately serve the citizens who generate it. Ethical integration of CGD into official statistics requires mechanisms that ensure communities benefit from insights derived from their contributions, whether through improved services, inclusive policies, or strengthened trust in institutions. This reciprocity transforms data from an extractive exercise into a collaborative social good.

We situate these ethical imperatives within Kenya's socio-political landscape, where citizen participation is often mediated by civil society organizations and grassroots networks. While these actors play a crucial role in mobilizing communities, their credibility and neutrality must be safeguarded to prevent data capture by partisan or self-serving agendas. The paper highlights case examples where CGD initiatives have succeeded in balancing statistical integrity with ethical responsibility, as well as scenarios where credibility may undermine trust.

In the paper, a framework for ethical CGD integration is proposed, premised on: consent, integrity, reciprocity, and credibility. By foregrounding these principles, Kenya's statistical system can harness the richness of CGD while protecting citizens' rights and ensuring that official statistics remain a trusted public good. This contribution extends debates on data ethics by demonstrating that ethical considerations are not peripheral but central to the legitimacy and sustainability of participatory data ecosystems.