Data & Policy (2026), 8: e16
doi:10.1017/dap.2026.10065

RESEARCH ARTICLE

Reading gender into citizen data practices for sustainable development

Tara Patricia Cookson[1] and Lorena Fuentes[2]

[1]School of Public Policy and Global Affairs, The University of British Columbia, Canada
[2]Ladysmith Collective, Canada
Corresponding author: Tara Patricia Cookson; Email: tara.cookson@ubc.ca

Received: 12 February 2025; **Revised:** 26 March 2026; **Accepted:** 10 April 2026

Keywords: citizen-generated data; data revolution; development; gender data; gender equality

Abstract

Citizen-generated data (CGD) is increasingly embraced as a strategy for filling data gaps to achieve the United Nations Sustainable Development Goals (SDGs), including SDG 5, Gender Equality and the Empowerment of Women and Girls. Existing frameworks to guide the design and use of CGD, however, do not reflect the unique considerations of CGD projects addressing issues of gender inequality. Answering recent calls for study of "data practices," this article analyzes common CGD principles through findings from an action research project to address gender-based violence at the Colombia–Venezuela border. We suggest that while existing frameworks provide generative pathways forward, several principles around which consensus appears to be emerging are too rigid or insufficiently nuanced to account for the dynamics that many gender-focused CGD projects confront. These include dynamics such as physical security, the role of emotion in shaping project implementation, unequal access to resources, and political will. We suggest that this rigidity truncates the utility of these frameworks for CGD actors navigating highly sensitive issues in risky environments where serious violations of women's human rights are taking place—and where the generation of gender data is one of several motivating factors for the work being done. This article reads gender into these frameworks to broaden the range of sustainable development issues for which CGD can help catalyze progress.

Policy Significance Statement

Significant gaps exist in the availability and quality of official national data on a range of issues related to gender equality and women's empowerment, particularly on pernicious issues like sexual and gender-based violence (SGBV) and in challenging contexts. This study provides the first modified framework for generating citizen data in risky environments where serious violations of women's human rights are taking place. In doing so, it fills a gap in emerging efforts by development and human rights organizations to standardize practices around citizen data for inclusive policymaking and resource allocation.

1. Introduction

Citizen-generated data (CGD) is increasingly embraced as a strategy for filling data gaps that inhibit the capacity to track progress on the United Nations Sustainable Development Goals. The Collaborative on Citizen Data, established in 2023 following a recommendation by a UN Expert Group, recently drafted the Copenhagen Framework on Citizen Data to advance a common understanding of citizen data and its

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

potential uses (Collaborative on Citizen Data, 2024). A growing list of over 430 CGD projects from around the world is recognized by a multilateral data-for-development movement, the Global Partnership for Sustainable Development Data (See: https://docs.google.com/spreadsheets/d/1KVjxqByUH6ZCfpKmX8q7qBJX96Q8K5P70NRseP0jWzU/edit?gid=0#gid=0). And it is not just UN actors jumping on board: the European Union, Finland, Germany, the Netherlands, and the United States (just prior to the USAID closure) launched Craf'd, an open call for CGD on gender-based violence in fragile and crisis contexts. While CGD has a somewhat more robust history of use in environmental monitoring and, to a lesser extent, good governance, the recent high-level interest in citizen-generated *gender* data is newer.

The interest in citizen-generated gender data is notable given significant gaps in official data sources on some key topics. According to UN Women, the global average of data availability across the 52-gender-specific Sustainable Development Goals indicators is only 56% (UN Women and DESA, 2024). Collecting data on some topics, such as sexual and gender-based violence (SGBV), is especially challenging due to the sensitive nature of the topic (del Villar et al., 2025). SGBV victims may be disincentivized from reporting violence because of perceived and actual impunity for perpetrators, because there is nowhere to report it, because violence is normalized, or because they may fear repercussions. Meanwhile, government actors may not want to collect data on SGBV if it casts them in a negative light. The Demographic and Household Survey has made significant contributions to generating population-based data on SGBV in countries around the world. Yet even this dataset, which is widely recognized as a trustworthy resource for data-driven decision-making, is insufficient for identifying key information about SGBV that practitioners and policymakers can use to make context-specific decisions (del Villar et al., 2025). This includes, for example, data on nuanced social, economic, and institutional drivers, what works to reduce violence, and real-time changes in risk and exposure to violence in times of shock or crisis.

Women's and feminist movements have long made use of unofficial evidence and data to draw attention to gendered inequalities and injustices, from time poverty to SGBV and feminicide (D'Ignazio, 2024). Yet these sources of data have not conventionally been accepted as legitimate sources of evidence. Recent advances, such as the Copenhagen Framework, suggest that the tide may be turning, with CGD initiatives being recognized as important for complementing official statistics and holding governments to account.

Despite these advances, existing research and guidance on citizen data for sustainable development are largely focused on a narrow set of development issues, with a large gap around the use of CGD to address issues of gender inequality and discrimination. A growing set of frameworks guides the production, dissemination, and use of CGD, often drawing on real CGD projects to propose a set of principles around important topics such as participation and inclusion, data privacy, and quality control, among others. Few of these, however, explicitly adopt a gender lens—in other words, approach the issue of generating gender data as something that may require unique considerations (cf. Uganda Bureau of Statistics, 2020).

Similarly, the existing scholarly literature that draws lessons from implementing CGD projects (Heeks and Shekhar, 2019; Balestrini et al., 2021) is largely focused on sustainable development issues where a gender lens is infrequently applied. This gap in the CGD literature is worth addressing because the drivers of inclusion and exclusion, and the risks and vulnerabilities associated with gender issues, are often distinct. A specific focus on citizen-generated *gender* data also offers an opportunity to address long-standing feminist calls for women's knowledge to be valued in development policy and practice.

This article fills the gap in the scholarly and practitioner literature on CGD and gender equality. It does so while simultaneously answering Lämmerhirt et al's recent call for studies of *data practices*: "the analysis of data uses and policies, as they are articulated, understood, or turned into situated activities by different actors in specific contexts, involving rules, socioeconomic factors, discourses, artifacts, and other elements" (2024, p. e24–3). In line with this call, our intention with this article is to guide the development of CGD practices that reflect the relationship between data, its intended uses, and how it comes into being (Lämmerhirt et al., 2024)—and, specifically, to guide the use and adaptation of CGD frameworks and policies to generate data on gendered issues.

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

The research questions guiding this article are as follows: (1) to what extent are current CGD frameworks attentive to gender?; and (2) how might they become more useful for projects oriented toward gender equality and gender justice issues? In pursuing these questions, our aim is to examine and improve the use of CGD as a tool for achieving gender equality and the empowerment of women and girls. To do so, we draw on a case study of a CGD project called *Cosas de Mujeres* ("Women's Stuff") that sought to generate data on gender-based violence in the context of a development-humanitarian crisis at the Colombian–Venezuela border.

Drawing on action research findings from this project, including document analysis and qualitative interviews, we engage with six common principles across CGD frameworks in the scholarly and practitioner literature, and for which we suggest that a gender perspective is lacking: participation/inclusion (and the role of engagement in fostering this), data privacy and security, data quality and interoperability, incentives, technology, and informed consent.

The frameworks reviewed provide generative pathways forward for CGD actors motivated to engage in data activism and advocacy across a myriad of development issues. Yet we suggest that the emerging consensus principles are often too rigid or insufficiently nuanced for CGD actors navigating highly sensitive issues in risky environments where serious violations of women's human rights are taking place —and where the generation of gender data is one of several motivating factors for the work being done. Instead, we emphasize considerations such as physical security, emotionally charged implementation contexts, unequal access to resources, and political will. While our case study, *Cosas de Mujeres* (CDM), focused primarily on gender-based violence, it touched on other critical gender equality issues, including sexual and reproductive health and rights and access to economic resources (see Cookson and Fuentes, 2026). Thus, the principles we propose in this article are relevant to a broader range of gender equality issues that CGD projects might tackle.

The article begins with a review of scholarly and grey literature on CGD, including a review of existing CGD frameworks. The next section introduces our methods, including our case, CDM, and the data sources used. Following this, we present our findings, discussing these in relation to the common principles identified across these frameworks. We conclude with recommendations for research and practice.

2. Literature review: Citizen-generated data

CGD is an action-oriented data source that involves community participation in collecting information on a problem that matters to them. It is "data that people or their organizations produce to directly monitor, demand or drive change…providing direct representations of their perspectives and an alternative to datasets collected by governments or international institutions" (Datashift, 2017, p. 5). In the context of the SDGs, it can complement the official data sources used to track progress on the goals, such as national statistics, many of which are incomplete. It can also be used to verify official reporting data and to generate "shadow reports"—civil society-led accounts of national progress in international monitoring processes —and strengthen accountability (Datashift, 2017, p. 9). CGD initiatives can be initiated, technically supported, and financed by governments, businesses, and international institutions, even as they rely on ordinary people ("citizens") and often civil society to function. CGD can be particularly useful to governments in the context of humanitarian and environmental disasters where information is needed quickly, as well as in highly specific contexts to solve local problems that are not captured in official statistics (Uganda Bureau of Statistics, 2020).

CGD overlaps in significant ways with citizen science, an approach to knowledge production that involves the participation of citizens throughout the execution of a research project, and which can but does not have to include specific policy-change aims (Fraisl et al., 2020; Balestrini et al., 2021). CGD can include quantitative data and big data or qualitative data and can make use of a range of different data sources and methods, which may include the use of technology in various ways, from crowdsourcing information from individuals through their mobile phones (Cookson and Fuentes, 2026) to environmental sensors (Gabrys et al., 2016; Mahajan et al., 2022).

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

Various conceptual frameworks and toolkits are emerging to guide CGD projects. At the highest level, the Copenhagen Framework on Citizen Data reflects the growing effort to produce a global consensus on what CGD is and how it can help advance the SDGs, including through collaborations between CSOs and state institutions (namely, National Statistics Offices). The Copenhagen Framework outlines 10 principles concerning the production and use of data by citizens for public policy and SDG monitoring. These include principles of participation and informed consent, data security, and data openness and accessibility, amongst others (Collaborative on Citizen Data, 2024).

Whereas the Copenhagen Framework provides a conceptual framework, typology and operational definition geared at generating buy-in around CGD from different stakeholders, not least national governments, other toolkits seek more explicitly to guide the design of CGD projects (Civicus, 2017; Datashift, 2017; Global Partnership for Sustainable Development Data, 2017; Uganda Bureau of Statistics, 2020; Kenya National Bureau of Statistics, 2023). At a high level, these frameworks seek to establish a standardized approach to generating CGD, typically with a view to producing "reliable, usable and accessible data" that may even be assessed for compliance with national data quality standards (Uganda Bureau of Statistics, 2020, p. vi). The guides prompt users to consider the intended audience, stakeholder incentives (benefits to participants, donors), and basic research methods questions, such as the type of data desired and how technology can be used to facilitate the data collection process. Some of these guides avoid the more prescriptive approaches that can accompany "top down" frameworks, and instead offer criteria to help civil society actors, as well as public institutions, policymakers, and funders navigate the opportunities and risks of "extra-official" data generated through CGD (see, e.g., Gray et al., 2017).

While CGD is a growing field of practice, it remains significantly under-studied in the peer-reviewed literature (c.f. Meijer and Potjer, 2018; Thuermer et al., 2024). Existing scholarship emphasizes CGD projects as technical *and* social, in that they involve software and networks as well as actors, institutions, relationships, and policies (Balestrini et al., 2021, p. 11). The various actors involved may seek and experience different benefits (Gabrys et al., 2016; Ponti and Craglia, 2020). While citizens may be seeking an immediate benefit from their involvement, governments, donors, or international organizations may be in it for the long haul, hoping to make eventual policy change (Lämmerhirt et al., 2016; Balestrini et al., 2021). CGD can be a tool for enabling ordinary people to exert their rights by generating evidence on issues that they care about (Gabrys et al., 2016), while for public sector institutions, CGD can offer timely and more granular data than available through official statistics, which may be updated only every few years (if at all, in some cases) (Lämmerhirt et al., 2019). Sustained citizen engagement around "matters of concern" is a key factor shaping the extent to which CGD projects can "scale, spread and sustain themselves" (Balestrini et al., 2021, p. 2).

Yet CGD and the technologies used to generate it do not inherently guarantee liberation. The uptake of CGD may be undercut, for example, by intended data users (decision-makers) viewing CGD with suspicion and thus rejecting the data on grounds of quality and reliability (Gabrys, 2022). The fitness-for-purpose, including interoperability and trustworthiness of the data produced, is a major concern in the CGD grey literature (Lämmerhirt et al., 2019). Discussing citizen data in the context of air pollution, Gabrys (2022) writes that "citizen data do not guarantee a remedy to the problems documented. Instead, data become a medium through which to figure worlds by monitoring, documenting, narrating, and analyzing conditions of disenfranchisement and dispossession" (Gabrys, 2022, p. 152). Thus, rather than CGD being comparable to official or conventional data sources, it may be "just good enough" to draw attention to an issue that ordinary people experience but that might otherwise be unknown or ignored by authorities (Gabrys et al., 2016).

The benefits that participating actors hope to accrue may not materialize equally. Some scholars caution that CGD "threatens to responsibilize individuals and communities" in data collection efforts, even where there is little hope that these efforts will lead to improvements in their lives (Crooks and Currie, 2021). Other scholars raise concerns that while CGD may lead to some gains for marginalized communities, they still risk increasing *relative* inequalities between groups with varying proximities to power (Heeks and Shekhar, 2019; Lewenstein, 2022). Hoped-for project benefits can also be unevenly

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

realized or not unrealized at all when incentives, prioritization, and ownership among participating actors are misaligned—outcomes that may be mitigated by adequate planning (Thuermer et al., 2024).

To date, very little of the CGD literature focuses on gender equality as a sustainable development issue (cf. McIlwaine et al., 2023). Perhaps because of the overlap with citizen science data, much of the literature concerning the contributions of CGD to development has focused on environment- and pollution-related issues, including targets related to SDG 6 (water and sanitation) and SDG 11 (urban development) (Fraisl et al., 2020; de Sherbinin et al., 2021). While scholar-activists like Catherine D'Ignazio and others have traced the design and implementation of "counterdata" or "data activist" projects focused on monitoring feminicide and gender-related violence at the grassroots and in some cases national level, they note that these initiatives are mostly delinked from international development frameworks and considerations (D'Ignazio et al., 2022, p. 25; see also Chenou and Cepeda Másmela, 2019). In contrast, this article explicitly tethers a case study of citizen-generated GBV data to the broader CGD frameworks emerging across the international development ecosystem. In doing so, it aims to interrogate the extent to which the emerging consensus principles accommodate the unique considerations that accompany gender-focused CGD projects.

3. Methods

This article presents a qualitative case study of CDM, a citizen-generated gender data project. Case studies fulfill a range of unique functions in social science research. Significant among these is the identification of causal relationships and "strategic structure," that is, "how interaction effects of one kind or another influence options, processes and outcomes" (Widner et al., 2022, p. 4). Case studies thus document how pre-existing contextual factors like environment or income influence outcomes, or similarly how practices of negotiation and contestation do (Widner et al., 2022, p. 8). For our project, using a case study enables us to ground a critical-constructive engagement with CGD frameworks and principles that exist "on paper" in the messy realities of "doing" gender and development work via the implementation of a gender data project.

We adopt an action research approach to this case study for two main reasons. First, while action research encompasses a range of different models, the approach is broadly distinguished by a dual commitment to address practical concerns and contribute to social scientific knowledge (Koshy, 2005). As "constructive enquiry" (Koshy, 2005, p. 9), action research is an attractive mode of knowledge production for us as researcher-practitioners (Holeman and Kane, 2020; Dekker et al., 2022). Our engagement with CDM at the time of its implementation was practice-oriented—we were operating as development practitioners—roles in which we continue to engage. However, we also brought and continue to bring our scholarly interests and training to bear on the project. This includes our knowledge of the literature on GBV, migration, and social protection, experience in the design of data collection tools, and a desire to contribute a gender perspective to the growing bodies of literature around data, technology, and development.

Second, action research involves "learning in and through action and reflection" (McNiff, 2013, p. 24), including by "thinking carefully about the circumstances you are in, how you got here, and why the situation is as it is" (McNiff, 2013, p. 25). Rather than studying a project as outsiders, we assumed lead roles in fundraising, context and user research, project design and iteration, research on technology considerations and requirements, network building at local, national, and international scales, responding to the WhatsApp messages, analyzing the data, and disseminating findings. These embedded experiences and roles permitted us to learn, to critique, *and* to construct.

Our data sources for this article draw on five sources of primary data in English and Spanish. This included 2 years of 1) field notes; 2) internal project communication (e.g., emails, project documentation); 3) program policies and protocols; and 4) external project presentations, all of which were generated while CDM was in its design phase, implementation, and closeout. We (article authors and research assistants) further conducted interviews with 14 former CDM staff members after the project was completed (our fifth data source). This included interviews with leadership staff, *gestoras digitales* (digital managers)

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

who operated the WhatsApp and Turn.io platforms and systematized the data, and *gestoras comunitarias* (community managers) who were responsible for community engagement and project socialization.

While CDM was operational, all team members contributed in some capacity to the analysis and/or dissemination of the data generated in the WhatsApp message threads. This includes primarily data about incidences of violence, requests for resources, and limited demographic data. This article does not draw on the data from these message threads. Rather, it draws on the data sources outlined above, which CDM staff generated through their routine interactions with migrants, residents, and service providers in the communities where we operated. We regularly discussed these data in team meetings and acted on it in various ways. For example, we published data briefs and op-eds, and hosted workshops and webinars to disseminate them to a range of audiences (e.g., government officials, funders, service providers). We also used what we were learning to iterate on the program, both in terms of what data we did (and refused to) to collect, as well as what services we provided and how (for further details, see Cookson and Fuentes, 2026).

In this article, we draw on these data sources as well as the 14 interviews we conducted as part of a post-CDM academic project. Taken together, these sources of data helped us as authors to reflect on the possibilities and limitations of existing CGD frameworks. All Spanish-language data presented in this article have been translated by the authors.

3.1. Cosas de Mujeres*: A citizen-generated gender data project*

CDM operated in Colombia between January 2020 and August 2022. At the time, Colombia had been receiving some 2.5 million migrants from Venezuela annually since 2015 in what became the largest recorded refugee crisis in the Americas (UNHCR, 2022, p 5). Colombia was also the site of conflict-fueled internal displacement of more than 8.5 million people, as recorded in the Government of Colombia's Victims Registry (Government of Colombia, n.d.), existing in what some refer to as a "humanitarian-development nexus" (Hinds, 2015). CDM operated in three cities: Cúcuta (pop. 711,715) on the border with Venezuela, Cartagena (pop. 914,552) on the Caribbean coast, and Bucaramanga (pop. 581,130), an inland city. The project was funded by USAID Colombia Transforma, Global Affairs Canada, UN Women, Turn.io, and Ladysmith.

The idea for CDM followed from the 2019 publication of an op-ed in the Washington Post written by one of the project's founders, Dr. Julia Zulver. Julia had been conducting fieldwork for an unrelated project when she began receiving reports of GBV and unmet need for prevention and response services. Upon questioning, local officials explained that while they too had observed the problem, they were unable to respond because of a lack of systematically collected data. In the words of one official, as a result, "perpetrators take advantage of the fact that [women] are almost invisible" (Zulver, 2019). While frontline humanitarian and development agency staff knew that GBV was widespread, they cited a lack of data as preventing them from acting. This article's authors had been working in the gender and development field and in "gender data" specifically. Upon reading the op-ed, they approached us (article authors) about developing a project that addressed the specific barrier to action cited by the service providers: a lack of gender data. As such, CDM shares a common CGD motivation: to raise awareness of a topic that receives insufficient attention from institutions (Ponti and Craglia, 2020).

3.2. Project design

We launched the project in Cúcuta after a design phase in which we consulted with intended data users, including program staff and lawyers at justice institutions, local women's rights advocates and civil society organizations serving women and LGBTQ populations, social workers, clergy at a Catholic church serving peripheral neighborhoods, local academics focused on GBV, NGO staff at clinics that provide women's health services, multilateral humanitarian and global development organizations, and foreign aid institutions. We also consulted migrant women (and some men, when they approached us) about their access to services and resources, including access to and use of technology (e.g., phones,

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

internet, data). Consultations took place in locations where we knew migrants were located: at the border crossing, at health clinics and dispersal points for in-kind transfers, and in a neighborhood where women engaged in sex work.

Through the design phase, we learned about various constraints service providers faced, in particular, a shortfall of resources to meet demand. We also learned about the deeply challenging context that migrant women navigated. After crossing into Colombia via pathways where SGBV was common, they arrived at a city and a system of social services that was entirely new to them. We found that they were often unaware of the services that did exist. Many arrived with children in tow, for whom they were solely responsible because their partners had migrated earlier and elsewhere. They were not eligible to work formally and, as such, scrambled to secure food and rent in overcrowded housing, often in quite peripheral neighborhoods. Women sold their hair and engaged in survival sex to make ends meet. Venezuelan women were framed in popular narratives as hypersexualized, a stereotype that informed the kinds of sexual violence perpetrated against them. At the same time, we were made aware of instances of Colombian women opening their doors to Venezuelan women in an act of solidarity. We also learned that many migrant women had access to a smartphone (either their own, or through a family or community member) and were using WhatsApp to coordinate their journeys and stay in touch with family located elsewhere (For example, a WhatsApp-based service "Estoy en la Frontera" (I'm at the border) was run by the local newspaper La Opinion. It provided migrants with information about health, education, legal, and employment opportunities to migrants and collected data on location and sex. We learned that 80% of the messages they received were from women. Migrants also used WhatsApp groups to coordinate group transportation to other countries. WhatsApp is currently used by other service providers and migrants in Latin America to facilitate access to services like sexual and reproductive healthcare (e.g., "Te Acompaño" in Mexico)).

Ultimately, we landed on a project that sought to address GBV through three pathways: by using WhatsApp to provide women with information about where they can access services that prevent and respond to GBV; by generating anonymous data that can be used to inform a medium and long-term advocacy strategy for more effective service delivery; and by fostering solidarity among Venezuelan and Colombian women in host communities (This included Colombian "retornadas," women who had fled Colombia for Venezuela as a result of civil conflict in Columbia since the 1980s). Our approach aimed to avoid a "data extractivism" model that collects personal data to generate value that may not be experienced as a benefit to the population from whom the data were collected (Horst et al., 2024). Rather, we sought to meet women's immediate needs by providing them with information about existing services, as well as in the medium- and long-term, by using the data to advocate for service improvements that would ultimately reduce vulnerability to violence and better respond when it was perpetrated. Furthermore, whereas extractivist approaches prioritize data generation and broader research or other objectives above the rights of beneficiary communities (Garcia et al., 2020), we pursued a data minimization approach. This was in part motivated by our desire to only collect data that could easily lend itself to policy action, as well as an adapted response to women withholding certain data that we initially did solicit (see Cookson and Fuentes, 2026).

Following advice from a local feminist organization, we advertised the project as one addressing broadly "cosas de mujeres" (women's stuff) to reduce the risk that women would be made unsafe if found to be interacting with an anti-violence project. Our team of local *gestoras comunitarias* (community managers) disseminated project information through GBV workshops, posting flyers with the WhatsApp number on telephone poles and bathroom stalls in bars, painting murals on buildings in common thoroughfares, speaking with people about it at service points such as health posts, beauty salons, and soup kitchens, hosting joint events with other civil society and government partners, and networking with community leaders.

3.3. Technology and data

We made use of an evolving set of digital tools throughout the project. Our first prototype used the generic WhatsApp platform available to any individual user, and we adopted various automations as the project

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

evolved. These automations were facilitated by a transition to WhatsApp Business and a partnership with Turn.io, a web-based platform that enabled us to have multiple *gestoras digitales* (digital managers or phone operators) responding to incoming messages as we scaled to multiple cities. In the project's final iteration, when someone messaged the WhatsApp number, they received an initial automated response explaining that CDM provided information about services relevant to women, asking for consent to share de-identified data, and asking how CDM could help. Afterwards, a CDM telephone operator took over the messaging. The phone operators included Colombian or Venezuelan women living in Colombia. The conversation from that point on was between two humans.

The content generated by these message threads was our data. From the message threads, and where women consented, we extracted anonymized quantitative and qualitative data on challenges women were facing and their perceptions of their own needs. We never asked for names, identification numbers, street addresses, or any other identifying information. The data points we collected and disseminated were determined through a process of "negotiated refusal" that involved contestation and agreement between our team members and prospective data users and donors, as well as among team members, about what data were useful and reasonable to collect (Cookson and Fuentes, 2026).

We routinely analyzed the data and disseminated them in the form of workshops and "Gender Data Briefs" with the intended data users we initially consulted, plus others in the consortium of international organizations working on the refugee and migrant crisis response (GIFFM), the local police who received and could respond to GBV reports, and the national statistics office (Departamento Administrativo Nacional de Estadística, 2021).

With the securing of additional donors, the project "spread" to Cartagena and Bucaramanga, where it operated until 2022. We used broadly the same approach to setting up the project infrastructure in these cities: starting with baseline field research that involved speaking with and partnering with local women's and feminist organizations, mapping the landscape of service providers, and hiring local *gestoras comunitarias.* Despite CDM "scaling" to three cities, the work remained focused on maximizing human-to-human interaction on the messaging platform and leveraging community engagement, which we understood from feedback from women themselves and our local staff was essential to the project's success in securing trust and buy-in from service users and local providers. Conversations with prospective donors regarding "scale" were more challenging (Balestrini et al., 2021). These conversations typically revolved around the adoption of additional forms of automation, often discussed in the language of "innovation." Our decision to push back on these requests—because we did not perceive such "innovations" to be appropriate in light of best practice on the safe and ethical use of technology to address GBV (see UNFPA, 2023) nor useful or necessary to meet our project aims, which donors themselves (and more importantly, CDM beneficiaries) said we were succeeding in meeting—may have contributed to waning donor interest and ultimately to the project funding ending.

CDM met the criteria for a CGD project by being explicitly problem-focused, by filling gaps in national statistics, and by involving communities in collecting data on a problem that affects them directly. Our case adds to the CGD literature with a case that is neither purely "bottom-up"—in the sense that the community self-organized—nor "top-down" in the sense of an initiative conceived of and coordinated by the highest levels of government or international organizations (Balestrini et al., 2021). Rather, through the varied positionalities and perspectives of those who implemented CDM, the case draws learnings "from the middle."

4. Reading gender into CGD frameworks

This section of the article engages with six common principles across CGD frameworks: 1) participation/inclusion (and the role of engagement in fostering this), 2) data privacy and security, 3) data quality and interoperability, 4) incentives, 5) technology, and 6) informed consent. We organize our results by these six common principles because our intention as scholar-practitioners is to shape common "data practices" with a view to making them more effective tools for achieving the realization of women's human rights.

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

Using CDM as a prism through which the particular dynamics of a CGD project focused on such rights are brought into view, we consider whether these principles hold up to empirical scrutiny.

We argue that while each principle holds value, as commonly interpreted, they are either too rigid or insufficiently nuanced to account for the risks and sensitivities, as well as motivations and power dynamics, that shape contexts where serious violations of women's rights are taking place. Considering these limitations, we offer ways forward for "reading" gender into each overarching principle.

4.1. Participation/inclusion/engagement

The Copenhagen Framework suggests that "All groups of interest should be involved, including those that are vulnerable and marginalized, and participation should be free, open, equitable, accessible, and transparent." Some variation on participation, inclusion, and engagement is present across CGD frameworks, which emphasize the positive aspects of these principles, and considering these, the need to ensure that no marginalized group is left out. Engagement is a prerequisite for participation and inclusion, and thus the effectiveness and sustainability of a CGD project (Balestrini et al., 2021). In the context of CDM, the community managers were responsible for engagement.

Our case suggests that the more nuanced approach of the Global Partnership (GPSDD) is more appropriate: "Participation can vary in breadth and depth," such that "CGD does not always have to be about maximising the breadth of participation. Rather, governments should ask what kind of participation is meaningful and useful for an initiative" (Global Partnership for Sustainable Development Data, 2017, p. 4). Our case illustrates that this is because engagement in feminist mobilizing and gender issues can entail risk (Zulver, 2022). The women most likely "left behind" and uncounted are those in the poorest and most underserved neighborhoods of the cities in which CDM operated. Due at least in part to a lack of responsive social, health, and law enforcement and judicial services, as well as an attendant context of impunity, they are also sometimes the places where violence is the most prolific. If we accept the premise that engagement is necessary for the effectiveness and sustainability of a CGD project, then the work of engagement must also be done in the areas where women and their experiences go uncounted. As described by one of the community managers: "So I mean, let's say that the budget, one understands that the budget of [these] institutions is often not enough. So then in the end, they are the most vulnerable, the most forgotten, and the least served. So that's where like, in general, we could say that we focused our interventions." (CDM Staff #12).

Doing the work of engagement in these communities to ensure that the most vulnerable women are included, however, has inherent risk. These were described at length by CDM community workers:

> We arrive at some parts, neighborhoods or settlements that are terrible. Terrible in the sense that there were shortages of everything, absolutely everything. The houses were ranches, there were no public services, there was a lot of drug addiction, a lot of prostitution. That is to say, it was a thing where one was like "uff, we arrived at a quite nerve-wracking place." And then getting there, starting with getting to those places, I had one, the transportation was, or well, I spent almost two and a half hours to get to the places…. and when we arrived it was like thank goodness. (CDM Staff #11)

> Well, there's a place where I did not return, called San Simón…because they immediately told me 'look girl, there are no leaders here, there is nothing here, everyone knows that whoever wants to work on sensitive issues of complaints, the Prosecutor's Office, etc, those who have dared to be community leaders or… They're gone or they're dead.' Then I say 'ah, well, too bad. Sure, perfect, I won't be back'. (CDM Staff #12)

Given its grounding in an imperative to ensure that "no one is left behind" and goes unrepresented in data or in the processes that go into collecting data, CGD frameworks tend to espouse the positive features of participation, inclusion, and engagement. However, our case suggests that the risks of participation and inclusion, including the engagement activities that are considered important for boosting participation and inclusion, require consideration for CGD frameworks to meaningfully incorporate a gender lens. This

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

could entail nuancing the principle so that those interested in designing and implementing a gender-focused CGD project are attentive to not only the potential *empowering* benefits of participation, inclusion, and engagement, but also to the trade-offs and risks for those generating data in a particular context. While our own case did not reveal risks associated with data dissemination, there will be other contexts in which disseminating CGD on sensitive issues of gender inequality and discrimination would be extremely dangerous for those involved. We did not, for example, actively collect data on sex trafficking even when some service providers requested it, due to the security concerns posed by organized crime for CDM staff in data collection as well as dissemination.

4.2. Data privacy and security

Risk is largely discussed in CGD frameworks and best practice literature as those associated with data privacy and security. Even some of the frameworks that integrate gender only consider cybersecurity (Uganda Bureau of Statistics, 2020). As discussed above, a rigid interpretation of inclusion and participation, and a too narrow interpretation of risk, overlooks the risks inherent to collecting data on violence and the strategies needed to make a CGD project "scaleable, spreadable and sustainable" (Balestrini et al., 2021).

Our case surfaces two ways in which CGD frameworks fall short from a gender perspective in how they address risk in relation to data privacy and security. First, there is a largely brushed-over tension between widespread calls that "inclusive" data should be disaggregated by as many (usually identity-related) variables as possible, and principles for data privacy, confidentiality, and security. We confronted this tension when our intended data users and donors pushed us to collect more data about the women messaging our platform. In part because we had some relative power to do so, we were able to push back on these demands for more granular data, a practice we refer to as "negotiated refusal" (Cookson and Fuentes, 2026). We did this by leveraging our team's collective knowledge and experience in technology, data, and GBV advocacy to make the case that the risks to data privacy *and* the safety of women messaging us outweighed any potential gains vis-à-vis the "richness" of the data:

> I think that was another challenge of donors and people asking for this data really wanted us to be able to say, well, how many migrant women are writing to you? How many sexual minorities are writing to you? How many, etcetera? And we had explained that we don't ask them for [all] that information…. But there was a lot of identity data points that we are not going to ask for directly… Having to educate those who are interested in our data on that and kind of push back like, well, "this is why we're not going to ask them ten different identity questions before giving them information or service they need right now (CDM Staff #1 Interview)

A second way in which the discussion of risk around data privacy and security falls short in the CGD frameworks is that it pays insufficient attention to the role that emotion and affect play in data practices. The reality is that "practice is messy and rarely adheres cleanly to pleasing principles" (D'Ignazio, 2024, p. 14). Our case shows how even where CGD projects use rigorous protocols designed to keep data, team members, and beneficiaries safe and secure, the involvement of humans necessarily means that emotions play a role, and things can go "off script"—perhaps particularly when the project addresses highly sensitive issues affecting vulnerable populations.

> With the *gestoras digitales*, one of the challenges is that women would write and say, 'can I talk on the phone instead?' And we had […] we tried to set up protocols like, 'okay, here's, like, this is when you can this is how we do it, this is how we do it. Make sure you protect their privacy' and things like that. And I could tell from conversations with the *gestores digitales* that the protocol isn't always perfectly followed. And I think this goes to the heart too of just the challenge of safely providing GBV services, safe for the team and safe for those who ask for services, right? If you're a human with an empathetic heart and you're hearing this person that's like, 'I'm in this awful situation, but I

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

> just, can I call you instead?' like our *gestoras digitales* wanted to talk with those women, right?… (CDM Staff #1)

Insofar as it is oriented toward placing guardrails around citizens' personal data, the principle of data privacy and security is, of course, paramount. Yet reading gender into this principle would entail both expanding it to incorporate considerations of when less rather than more data collection is appropriate (Cookson and Fuentes, 2026) and nuancing it to acknowledge the role that human emotion and affect play in data practices and the potential impact on those doing the data collection (D'Ignazio et al., 2022). Both are aided by listening to "feedback" from the people doing the data collection and iterating accordingly.

4.3. Data quality

Data quality tends to be raised in two ways. One is in terms of having rigorous methods that can help to reassure the end user that it is trustworthy. This is particularly the case for qualitative data, which tends to be considered less objective and therefore less rigorous in regimes of evidence-based decision-making that drive international development and public policy (Eyben et al., 2015). The other is in terms of interoperability, which entails the capacity to use CGD data with existing and used datasets, for purposes of comparison and complementarity. In both cases, the imagined end user tends to be a government actor who otherwise relies on official statistics to make decisions.

Not all CGD data will be used by government actors. This may be because the aims of the data initiative may be to catalyze change from a different actor, or because the motivations for data collection exist outside the bounds of a dialogue or relationship with the state (Lucchesi, 2019). It may also be because the government may lack the will or resources to act on the data (Crooks and Currie, 2021). In such cases, interoperability may not be a necessary goal, and efforts to meet internationally recognized data standards may be fruitfully supplanted by producing "just good enough data" (Gabrys et al., 2016) for "small data uses" (Heeks and Shekhar, 2019). This is a lesson that we learned organically and by being open to iteration. Instead of merely collecting data on incidences of violence or about the women experiencing the violence, we began to collect data on the service offering ("service mapping"). We considered this a kind of "ground-truthing" of the availability of services as cited by government, civil society, and international services (Cookson and Fuentes, 2026). In responding to an interviewer's question about monitoring service quality, one of our staff members reflected:

> Yes, it seemed very interesting to me because from my vantage point I saw it as a opportunity for following up, although it was not so much a premeditated objective of Cosas de Mujeres to follow up… This service mapping not only works as a directory where I deposit information [about existing services], but also as an instrument that allows me to monitor both the relationship and what is happening with, how the institutions are working for the population. So we said well, let's try to see what is changing over time according to what women tell us. But it was not something that Cosas de Mujeres [leadership] told us to do at the beginning, like 'we are going to monitor institutions and their operations… (CDM staff #10)

We disseminated the data that we collected on gaps in the service offering to relevant service providers. In some—though certainly not all—cases, the service provider was able to make modifications. In this instance, the data were "just good enough" to make a material improvement for women seeking immediate help.

> So because we already had connections with the organizations, we managed to tell them through our monthly reports, 'well, the women know that the services exist, but they cannot access them. Why? Because they are left very far'. Those who have children in daycare have to be at home at four in the afternoon because the children who get a place in daycare have to be picked up. But a woman is going to come here to the city center, often for a whole day, without being attended to. So they prefer not to use the services, because first it is far away for me, second it is a day when I am not going to

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

> earn any money for my household, so I prefer not to go to the services [….] And when we begin to make this articulation such that they really *understand* these women's reality, they won't deliver these programs from some office. Instead, they'll go further to where women and their needs are. (CDM Staff #9).

This implies that the power of CGD does not only lie in its adherence to the statistical standards of the dominant measurement regime nor in its interoperability with state and international systems. Other forms of data could be used to make tangible change. This includes richly descriptive and context-specific qualitative data that go beyond capturing the demographic characteristics of marginalized groups and the incidences of their problems, but that points to solutions and ways forward (see Cookson and Fuentes, 2026). It is also more honest in the recognition that state actors may not always be the ones most willing to acknowledge and act on an issue of gendered inequality and oppression. This observation challenges the premise of desired monitoring of a problem like GBV. In addition to issues that may indeed be of interest to the state (and these certainly do exist), for some issues, the most impactful data user may be direct service providers and civil society organizations engaged in shadow monitoring (Datashift, 2017; D'Ignazio, 2024).

4.4. Incentives

The reasonable question of why people participate in CGD projects is often discussed in terms of volunteerism, though there is some acknowledgement that some participants will be paid, depending on their role in the project. In most cases, it is assumed that the citizen will participate in data generation because of the benefit they will receive when the data collected elucidates a problem and contributes to its resolution. This logic requires some acceptance that the problem in question will not be immediately resolved. The case of GBV, however, troubles this assumption. GBV is notoriously hard to measure using conventional methods, and under-reporting is a problem in stable contexts, and significantly worse in contexts of political instability and crisis, in contexts of human mobility, and in underserved contexts where poor service provision undercuts opportunities to report (True, 2020). These dynamics are a challenge for CGD projects that aim to make visible and resolve problems of violence. Our case showed how the promise of a resolved problem in the medium-to long-term future functioned as an incentive for different involved groups.

First, most of the CDM community managers and digital managers had a range of incentives, including a salary, skills building and knowledge development, and network development, including with powerful authorities in their cities and states. They also acknowledged the potential and real value of the data produced:

> Gender violence is a statistic that is often invisible, made invisible, like a ghostly statistic in this country. So the fact that we can collect data and make a kind of diagnosis based on the work we do is very valuable…it gives tools to organizations, to civil society, to measure the problem a little and also the type of response that is being given… (CDM Staff #8)

Common across our interviews, however, was an additional consideration that complicates typical interpretations of incentives. These were the affective elements of being involved in the project, which were both positive and negative, and could be understood as incentives and disincentives for participating:

> Well, you could say that I'm always trying to protect myself too, but being human is like that–it's feeling, say, touched by some situations and trying to handle it professionally. I had previous experiences with children who had been victims of sexual abuse or exploitation. Situations in which I also learned to put a certain distance… to remove a little bit of guilt, because sometimes you feel like if you don't manage to give everything to that person, you feel like you didn't do what you had to do, maybe. (CDM Staff# 5)

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

Second, for a woman immediately experiencing violence, or whose children are experiencing violence, the promise of a resolved problem in the medium- to long-term future was unlikely to be incentive enough. We resolved to overcome this challenge by offering to meet an immediate need: for information about existing help.

> There's all these debates around extractive data practices where you just kind of, like, take data from people […] And so we were kind of aware of some of these debates and wanted to think about what we could also offer women that was in a little bit more of a short term capacity. So, say, even though we were collecting data that actually was intended to benefit the people whom we were collecting the data from, it still was kind of like a medium or long term thing, because it's like, then the data needs to get disseminated, and there needs to be uptake, and you have to convince people with the data to do stuff, and then there's a funding cycle, and then maybe a shelter opens or a policy changes. But we were like, what can we offer people immediately? […] So, for example, we were once in a kind of, like, red light district or a district where there was a lot of prostitution and sex work, and women were really, like, they needed attention around some sexual health and reproductive health stuff, and they didn't know that the Profamilia clinic that offered free services was, like, three blocks away or whatever it was. It was, like, within walking distance. And so it was like, there was very obviously, like, sure, we know there were some gaps in services, but there was some stuff, but there was that kind of, like, information gap. So we thought, oh, well, maybe we can provide information to women about this. (CDM Staff #7)

A third group for which incentives can be discussed is potential data users. While there is some discussion across these emerging CGD frameworks about incentives for this group, including state actors and service providers, much of it assumes that the primary (or even sole) barrier to data uptake amongst decision-makers relates to data quality or dissemination. While this can certainly be a barrier, CGD projects involve an entire ecosystem of actors with varying proximities to power and levels of interest in generating or making use of data that sheds light on inequalities and injustices. In some cases, CGD projects may elucidate dynamics that are already well-known to the state (Crooks and Currie, 2021). This is well-documented in relation to GBV and feminicide, particularly when data shows violence perpetrated against women additionally marginalized on account of being poor, of irregular legal status, racialized, or living in rural and peripheral settings (Fuentes, 2020; D'Ignazio et al., 2022). In such cases, data may never be "good enough," because data are not the critical barrier to action; instead, it may be political will and prioritization, or empathy with affected populations, that stands in the way.

It is here, in particular, that the principle around incentives would benefit from a gender lens. First, it could be expanded to encourage CGD projects to be sensitive to the distinct and at times contradictory motivations that shape interest in and engagement with a project and its data. And second, recognizing that the state or public authorities may not in fact use the data irrespective of its relevance or quality, it could provide parameters for CGD projects to go further than setting medium and long-term goals around data collection, dissemination, and uptake to consider short-term needs that the project could help meet.

4.5. Technology

Much of the CGD literature generated has roots in environmental concerns and thus involves discussion of technological tools such as environmental sensors that monitor air quality and water pollution, and geo-spatial tools. These technologies are discussed in terms of gathering interoperable data, more granular data, more efficient data collection, and offering opportunities for upskilling of citizens participating. Our case suggests that additional considerations are required in contexts where groups involved are severely under-resourced. In such cases, technology may be both an opportunity and a constraint. Our initial project design research revealed that many, though not all, women had personal or borrowed access to a smartphone, that WhatsApp was widely used, and that women did not always have data but could use

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

WhatsApp for free (no data needed). Our leadership team did background research on possible uses of WhatsApp for unconventional data collection and learned from other projects what was possible.

> …there's a real tendency for people to want to say, oh, there's an app for that, and we'll build our own app to do something, and there's a lot of friction to do that. You have to be able to install the app, use a new app. It's a real pain in the butt. And with WhatsApp everyone already uses it, and everyone knows how to use it, it's a very natural kind of interaction. So especially in Latin America, most businesses will have a WhatsApp number that you talk to, that's you talk to the business. It's just a very familiar model, and it made sense to leverage that model. And also it has added benefits of, like, it's end to end encrypted, which is a nice additional feature, that sort of thing. It just fits very well by having no friction and good security guarantees. (CDM Staff #6)

It was quickly apparent that WhatsApp would be a logical tool to use. This approach meant meeting communities where they were in terms of technology access and use. Even using such a basic technology offered some opportunities for upskilling when we switched from inputting data into an Excel sheet to using the Turn.io platform.

It is important to meet communities where they are in terms of technology and avoid the introduction of new tools unless the benefits are obvious and worth it (Balestrini et al., 2021). Even when defaulting to the most basic of technologies, some citizens will still be missed—and more often than not, those missed are more likely to be women, poorer, and living outside urban centers. As noted earlier, not everyone had access to a phone, and moreover, our platform did not permit the use of voice notes, which meant the *gestoras digitales* sometimes took additional and "off-script" steps to communicate with those women. This indicates the importance of a human and social element in CGD projects (and once again illustrates the affective and emotional, rather than merely technical, nature of the work):

> So many of [the women] they had WhatsApp and they used their phones to communicate with Cosas de Mujeres to receive orientation. One issue that we identified and that many times we also commented on, is that as some of them had Whatsapp, others did not even have a phone. So, what they did was, what we tried to do, was work through the [volunteer] community leader, as if the leader was, you could say, a source of connection with these platforms. So you don't even have a telephone, so you go to the leader and through the leader's telephone, they try to get in touch with you. (CDM Staff #10)

With some exceptions (Uganda Bureau of Statistics, 2020), the discussions around "technology" in the CGD toolkits and frameworks focus on the upside of technology, and do not explicitly engage with its gendered dimensions, including limitations. Reading gender into this principle could mean simply making clear that CGD projects are not "better" if they use more technology or introduce new technical tools. Our experience with CDM was clarifying on two main fronts. First, there is nothing wrong with pursuing the path of least resistance: any technology that is used in CGD projects needs to be appropriate for the context and population (rather than say responding to the desires of external actors for technocratic fixes or "innovation," e.g.). And second, our experience underscored the limits of technology and the importance of human-to-human engagement, particularly where a CGD project is collecting data on a sensitive issue with highly vulnerable populations.

4.6. Informed consent

The principle of informed consent holds that individuals should "provide informed and voluntary consent before their data are collected" and that "they should have control over their data and understand the purpose and use of the data collected" (Collaborative on Citizen Data, 2024). In theory, this is highly sensible, though in practice it can be much more complicated. It requires a lot of thought and is not simple to do effectively. We put a lot of effort into meeting the (still nascent) scholarly and human-rights standards that emanate in elite circles where this gets thought about a lot. Drawing upon our prior

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

experience in frontline GBV practice and thus our familiarity with GBV research and practice principles concerning informed consent, we experimented with changing the location of the consent message, and generating two versions, a short link in very plain language with an option to read the full policy.

> .. I think to a lot of people working on the ground, using the platform, running the platform, it doesn't feel as tangible. Especially we're talking about like gender-based violence, so it doesn't seem as real of an issue to invest time and resources in it. So I think that was also one of the struggles was like, you know, how far do we play with this? What is the success rate of asking for informed consent? [.....] If we were to hold participatory research workshops, would it be like to invest that time and resources in workshopping potential entry points to improve services? Or use these human resources to workshop what's a better way to securely collect data on this? If you were to ask social movement leaders, they definitely would have gone with the first rather than the latter. (CDM Staff #1)

In practice though, women who messaged CDM were often reluctant to or unclear about whether they wanted to provide data.

> …If the woman was in an urgent situation, she could ignore [the informed consent policy] and continue telling us her case. I'd say that [the policy] could be a little more limiting with people who don't, can't read or don't, didn't have like the understanding that in that link what they saw was more information about what we do with the information, something confusing, tangled… Many women thought that in order to access the information [about services], maybe I am wrong, but what I perceive is that many women thought that by entering that link they could receive the needed help, and they would tell us 'friend, I only have data for WhatsApp, I do not have data [to follow that link], help me here'. (CDM Staff #5)

We ultimately did not over the course of 2 years determine a fool-proof way of gaining informed consent and retaining a high acceptance rate. This was not because the women writing in were refusing consent based on a grounding in digital justice. As a result, the policy sometimes scared them off and generated less trust. What this dynamic suggests is a need for further work by digital justice and data rights communities to determine appropriate standards for informed consent when the trade-off is not generating data on the experiences of the most vulnerable groups. As with the principles around engagement and data privacy, the principle of informed consent would benefit from integrating insights from long-established principles from the GBV field (see Ellsberg and Heise, 2005) as well as greater nuance about where risk emanates from (e.g., our case shows that the risks related to community engagement may very well be more material in the short-term than risks around informed consent).

5. Conclusion

CGD is increasingly embraced as a strategy for filling data gaps that inhibit capacity to make progress on the SDGs, including most recently SDG 5. Existing frameworks to guide the design and use of CGD, however, could go further to address the unique considerations at play in CGD projects addressing issues of gendered inequality and discrimination. Through a study of "data practices" in a CGD project to address SGBV at the Colombia–Venezuela border, we distilled six principles common to "best practice" CGD frameworks that could be improved to better guide the development and implementation of CGD projects addressing issues of gender inequality and discrimination.

We suggested that while *participation and inclusion* are "development buzzwords and fuzzwords" that can make funders and practitioners feel good and may prompt well-intended action (Cornwall, 2007), critical thinking is required by all involved to weigh the potentially empowering benefits of involvement in data collection and dissemination against tradeoffs and dangers. The broadest possible achievement of participation may not always be feasible or even desirable, especially in highly insecure contexts or when defenders of women's rights are especially at risk.

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

Our case also motivates expanded engagement with *data privacy and security.* We presented a tension between "inclusive data" that demands ever more identity-related data points, and the potential risks this poses for identification of marginalized individuals and groups by nefarious actors. CGD frameworks might recognize that sometimes *less* data are preferable. We further acknowledged the challenges facing stringent application of privacy and security protocols when the individuals collecting the data are compelled by human suffering and emotions like empathy to act in ways that transgress project rules.

In relation to how CGD frameworks approach *incentives*—why people participate—our case showed that motivations of different involved groups can be distinct (Lämmerhirt et al., 2016). It additionally showed that participants may experience both incentives *and* disincentives, such as when the data they collect surfaces violations of rights that cannot be solved immediately, giving rise to feelings of helplessness, guilt and a desire to create distance from the problem. When the medium- to long-term promise of data as a tool for change is not incentive enough, CGD projects may consider including a project component that meets a short-term need. At the same time, we also suggested that further work by digital justice and data rights communities is needed to determine appropriate standards for *informed consent*, particularly when cumbersome consent, privacy, and security policies function as a disincentive for participation.

We also suggested that how *data quality* is conceptualized should go beyond achieving interoperability with national statistical and administrative systems (though this may in many cases be highly appropriate!), to also encompass data usability by a wider range of actors who can solve the problem at hand.

Finally, in addition to encouraging the use of *technologies* in data collection, we suggested a more nuanced approach that grapples critically with donor-driven requests for innovation considering suitability and necessity in the chosen context. We further called for the *human* elements of CGD projects to be emphasized and supported, particularly in cases where the physical safety and even lives of women and their dependents may be at risk. The loss of women's lives in AI-facilitated GBV projects should serve as important lessons.

Our case study focused on a CGD project that sought to address SGBV primarily. This case has the benefit of illustrating the outer bounds of considerations such as security risk and emotional distress, while also shedding light on other, even more moderate cases of inequality and discrimination. Yet there are many other issues that CGD could help address and for which empirically grounded case studies are needed, like time use or access to infrastructure, that may surface other important considerations. Practically, we hope our reflections prompt further discussion about CGD best practices in policy circles, with multilateral organizations, NSOs, and feminist and women's rights groups who use data to call attention to gendered inequalities and discrimination, and to see these addressed.

Data availability statement. The datasets generated during the current study form part of a broader research project examining an initiative responding to gender-based violence in a humanitarian crisis. Due to the sensitive nature of this topic and vulnerabilities of those involved, the datasets are not publicly available.

Acknowledgments. We would like to thank our Research Assistants, Lucía Mesa Vélez and Norma Patiño Sánchez, for their support with data collection.

Author contribution. Conceptualization: T.P.C., L.F.; Data curation: T.P.C., L.F.; Formal analysis: T.P.C., L.F.; Funding acquisition: T.P.C., Investigation: T.P.C., L.F.; Methodology: T.P.C., L.F.; Project administration: T.P.C., Resources: T.P.C., Writing—original draft: T.P.C., L.F.; Writing—review and editing: T.P.C., L.F.

Funding statement. This work was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) Insight Development Grant, and the Canada Research Chair program.

Competing interests. T.P.C. and L.F. are co-founders of Ladysmith, the organization that conceived of and operated the initiative featured in this study.

References

Balestrini M, **Kotsev A**, **Ponti M and Schade S** (2021) Collaboration matters: Capacity building, up-scaling, spreading, and sustainability in citizen-generated data projects. *Humanities and Social Sciences Communications 8*(169), 1–15. https://doi.org/10.1057/s41599-021-00851-5.

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

Chenou JM and Cepeda Másmela C (2019) NiUnaMenos: Data activism from the global south. *Television & New Media 20*(1), 396–411.

Civicus (2017) *Guide for Planning an Effective Citizen-Generated Data (CGD) Campaign*. Johannesburg: Civicus, Open Institute and Considr. Available at https://www.civicus.org/documents/toolkits/Guide-for-planning-an-effective-Citizen-Generated-Data-(CGD)-campaign.pdf (accessed October 15, 2025).

Collaborative on Citizen Data (2024) *The Copenhagen Framework on Citizen Data (v1). Background Document*. 55th Session of the United Nations Statistical Commission. Available at https://unstats.un.org/UNSDWebsite/statcom/session_55/documents/BG-4c-CGD_Framework-E.pdf (accessed October 15, 2025).

Cookson, T. P., & **Fuentes, L.** (2026). Gender Data, Intersectionality, and a Feminist Politics of "Negotiated Refusal". *Violence Against Women*, *32*(2), 347–373. https://doi.org/10.1177/10778012241309362.

Cornwall A (2007) Buzzwords and fuzzwords: Deconstructing development discourse. *Development in Practice 17*(4–5), 471–484. https://doi.org/10.1080/09614520701469302.

Crooks R and Currie M (2021) Numbers will not save us: Agonistic data practices. *Information Society 37*(4), 201–213. https://doi.org/10.1080/01972243.2021.1920081.

D'Ignazio C (2024) *Counting Feminicide: Data Feminism in Action*. Cambridge, MA: MIT Press.

D'Ignazio C, **Cruxên I**, **Suárez Val H**, **Martinez Cuba A**, **García-Montes M**, **Fumega S**, **Suresh H and So W** (2022) Feminicide and counterdata production: Activist efforts to monitor and challenge gender-related violence. *Patterns 3*(7), 1–14. https://doi.org/10.1016/j.patter.2022.100530.

Datashift (2017) *Using Citizen-Generated Data to Monitor the SDGS: A Tool for the GPSDD Data Revolution Roadmaps Toolkit*. Johannesburg: Civicus. Available at https://www.data4sdgs.org/sites/default/files/2017-09/Making%20Use%20of%20Citizen-Generated%20Data%20-%20Data4SDGs%20Toolbox%20Module.pdf (accessed October 15, 2025).

Dekker R, **Koot P**, **Birbil SI and van Embden Andres M** (2022) Co-designing algorithms for governance: Ensuring responsible and accountable algorithmic management of refugee camp supplies. *Big Data & Society 9*(1), 1–15. https://doi.org/10.1177/20539517221087855.

Departamento Administrativo Nacional de Estadística (DANE). (2021, July). Población migrante venezolana en Colombia: Un panorama con enfoque de género (Nota estadística). https://www.dane.gov.co/files/investigaciones/notas-estadisticas/jul-2021-nota-estadistica-poblacion-migrante-venezolana-panorama-con-enfoque-de-genero.pdf

Ellsberg M and Heise L (2005) *Researching Violence against Women: A Practical Guide for Researchers and Activists*. Washington, DC: World Health Organization.

Eyben R, **Guijt I**, **Roche C and Shutt CEyben R**, **Guijt I**, **Roche C and Shutt C** (eds) (2015) *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* Rugby: Practical Action Publishing Ltd.

Fraisl D, **Campbell J**, **See L**, **Wehn U**, **Wardlaw J**, **Gold M**, **Moorthy I**, **Arias R**, **Piera J**, **Oliver JL**, **Masó J**, **Penker M and Fritz S** (2020) Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science 15*(6), 1735–1751. https://doi.org/10.1007/s11625-020-00833-7.

Fuentes L (2020) "The garbage of society": Disposable women and the socio-spatial scripts of femicide in Guatemala. *Antipode 52*(6), 1667–1687. https://doi.org/10.1111/anti.12669.

Gabrys J (2022) *Citizens of Worlds: Open-Air Toolkits for Environmental Struggle*. Minneapolis, MN: University of Minnesota Press.

Gabrys J, **Pritchard H and Barratt B** (2016) Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data and Society 3*(2). https://doi.org/10.1177/2053951716679677.

Garcia P, *et al.* (2020) No: Critical refusal as feminist data practice. In CSCW '20 Companion: Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing, pp. 199–202.

Global Partnership for Sustainable Development Data (2017) Choosing and engaging with citizen generated data. Global Partnership for Sustainable Development Data, Open Knowledge International, Public Data Lab. Available at https://www.data4sdgs.org/resources/choosing-and-engaging-citizen-generated-data-guide (accessed October 15, 2025).

Government of Colombia. (n.d.). Registro Único de Víctimas (RUV). Unidad para la Atención y Reparación Integral a las Víctimas. Retrieved April 25, 2026, from https://www.unidadvictimas.gov.co/registro-unico-de-victimas-ruv/

Gray J, **Lämmerhirt D and Bounegru L** (2017) *Changing What Counts: How can Citizen-Generated and Civil Society Data Be Used as an Advocacy Tool to Change Official Data Collection?* Johannesburg: Datashift and Open Knowledge. Available at https://thedatashift.civicus.org/wp-content/uploads/2016/03/changing-what-counts-2.pdf (accessed October 15, 2025).

Heeks R and Shekhar S (2019) Datafication, development and marginalised urban communities: An applied data justice framework. *Information Communication and Society 22*(7), 992–1011. https://doi.org/10.1080/1369118X.2019.1599039.

Hinds R (2015) *Relationship between humanitarian and development aid.* Helpdesk research report relationship between humanitarian and development aid. United Kingdom Department for International Development. Available at https://reliefweb.int/report/world/helpdesk-research-report-relationship-between-humanitarian-and-development-aid

Holeman I and Kane D (2020) Human-centered design for global health equity. *Information Technology for Development 26*(3), 477–505. https://doi.org/10.1080/02681102.2019.1667289.

Horst HA, **Sargent A and Gaspard L** (2024) Beyond extraction: Data strategies from the global south. *New Media & Society 26*(3), 1366–1383. https://doi.org/10.1177/14614448231201651.

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press

Kenya National Bureau of Statistics (2023) *Validation of Performance Monitoring for Action (PMA) Kenya Data for Use in Official Reporting*. Nairobi: Kenya National Bureau of Statistics. Available at https://www.knbs.or.ke/reports/performance-monitoring-for-action-data-validation-brief/

Koshy V (2005) *Action Research for Improving Practice: A Practical Guide*. London: Paul Chapman Publishing.

Lämmerhirt D, **Jameson S and Prasetyo E** (2016) *Making Citizen-Generated Data Work: Towards a Framework Strengthening Collaborations between Citizens, Civil Society Organizations, and Others*. Johannesburg: Datashift and Open Knowledge International. Available at https://thedatashift.civicus.org/wp-content/uploads/2017/03/Making-Citizen-Generated-Data-Work_short-report_.pdf (accessed October 15, 2025).

Lämmerhirt D, **Gray J**, **Venturini T and Meunier A** (2019) Advancing Sustainability Together? Citizen-Generated Data and the Sustainable Development Goals. Available at https://www.data4sdgs.org/sites/default/files/services_files/Advancing%20Sustainability%20Together%20CGD%20Report_0.pdf (accessed October 15, 2025).

Lämmerhirt D, **Micheli M and Schade S** (2024) Exploring the practices of "data-driven innovation" in the European public sector. *Data & Policy 6*, e24.

Lewenstein BV (2022) Is citizen science a remedy for inequality? *The Annals of the American Academy of Political and Social Science 700*(1), 183–194. https://doi.org/10.1177/00027162221092697.

Lucchesi AH (2019) Mapping violence against indigenous women and girls: Beyond colonizing data and mapping practices. *ACME: An International Journal for Critical Geographies 21*(4), 389–398.

Mahajan S, **Chung MK**, **Martinez J**, **Olaya Y**, **Helbing D and Chen LJ** (2022) Translating citizen-generated air quality data into evidence for shaping policy. *Humanities and Social Sciences Communications 9*(1), 1–18.

McIlwaine C, **Ansari MR**, **Leal JG**, **Vieira F and Santos JS** (2023) Countermapping SDG 5 to address violence against women and girls in the favelas of Maré. *Rio de Janeiro, Brazil. Journal of Maps 19*(1). https://doi.org/10.1080/17445647.2023.2178343.

McNiff J (2013) *Action Research: Principles and Practice*. Abingdon: Routledge.

Meijer A and Potjer S (2018) Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly 35*(4), 613–621. https://doi.org/10.1016/j.giq.2018.10.004.

Ponti M and Craglia M (2020) Citizen-generated data for public policy: A brief review of European citizen-generated data projects. *JRC Technical Report.* European Commission, Ispra. Available at https://ec.europa.eu/jrc (accessed October 15, 2025).

de Sherbinin A, **Bowser A**, **Chuang TR**, **Cooper C**, **Danielsen F**, **Edmunds R**, **Elias P**, **Faustman E**, **Hultquist C**, **Mondardini R**, **Popescu I**, **Shonowo A and Sivakumar K** (2021) The critical importance of citizen science data. *Frontiers in Climate 3*, 1–7. https://doi.org/10.3389/fclim.2021.650760.

Thuermer G, **Walker J**, **Simperl E and Carr L** (2024) When data meets citizens: An investigation of citizen engagement in data-driven innovation programmes. *Data & Policy 6*, e12.

True J (2020) *Violence against Women: What Everyone Needs to Know*. Oxford: Oxford University Press.

Uganda Bureau of Statistics (2020) *Citizen Generated Data Toolkit*. Kampala: Uganda Bureau of Statistics. Available at https://www.ubos.org/wp-content/uploads/publications/09_2022UgandaCGD_Guideline_Toolkit_2910202102.pdf (accessed October 15, 2025).

UN Women and DESA (2024) *Progress on the Sustainable Development Goals: The Gender Snapshot 2024*. New York: UN Women.

UNFPA (2023) *Guidance on the Safe and Ethical Use of Technology to Address Gender-Based Violence and Harmful Practices: Implementation Summary*. New York: UNFPA.

UNHCR (2022) *Annual Results Report 2022-Colombia*. United Nations Refugee Agency. Available at https://reporting.unhcr.org/files/2023-06/2%20AME%20-%20Colombia.pdf (accessed October 15, 2025).

del Villar Z, **Spinardi A**, **Yáñez I**, **Sotolongo A**, **Lana AD**, **Fernandez Nieto B**, **Antoniassi Y**, **Fuentes-Nieva R and Letouzé E** (2025) Reflections on the relevance of updated, standardised, and reliable data on sexual and gender-based violence for research and policymaking for eradication, response, and prevention in Sierra Leone. *Data & Policy 7*, e14.

Widner J, **Woolcock M and Ortega Nieto D** (2022) *The Case for Case Studies: Methods and Applications in International Development*. Cambridge: Cambridge University Press.

Zulver J (2019) *Women are suffering extreme violence at the Venezuela–Colombia border, and the world isn't listening. The Washington Post*, February 26. Available at https://www.washingtonpost.com/gender-identity/women-are-suffering-extreme-violence-at-the-venezuelan-colombian-border-and-the-world-isnt-listening/ (accessed October 15, 2025).

Zulver J (2022) *High Risk Feminism in Colombia: Women's Mobilization in Violent Contexts*. New Brunswick, NJ: Rutgers University Press.

Cite this article: Cookson TP and Fuentes L (2026). Reading gender into citizen data practices for sustainable development. *Data & Policy*, 8: e16. doi:10.1017/dap.2026.10065

https://doi.org/10.1017/dap.2026.10065 Published online by Cambridge University Press