



# *To be Counted When They Count You: Words of Caution for the Gender Data Revolution*

April 2021

Author: **Noopur Raval**<sup>1</sup>

Editor: **Sumandro Chattapadhyay**

This study was undertaken as part of the **Big Data for Development** network supported by the **International Development Research Centre**, Canada, and is shared under **Creative Commons Attribution 4.0 International** license.

The Centre for Internet and Society, India

<https://cis-india.org>

---

<sup>1</sup> Noopur Raval is currently a postdoctoral researcher at the AI Now Institute at New York University. She can be contacted at [nar8914@nyu.edu](mailto:nar8914@nyu.edu).

# Table of contents

[Table of contents](#)

[Introduction](#)

[Background: Sustainable Development Goals and Governance by Numbers](#)

[Gendered Development and the role of data](#)

[Big Data for Gender Development: the new 'Gender Data Revolution'](#)

[Case Studies from the 'Gender Data Revolution' – using geospatial, digital exhaust and social media data](#)

[Broader Critical Reflection](#)

[Discussion](#)

[7.1. Big Data Solipsism in Welfare](#)

[7.2. The Uneven Geographies of Datafied Development](#)

[Conclusion](#)

[References](#)

# 1. Introduction

In 2015, after the announcement of the SDGs or Sustainable Development Goals, a new global developmental framework through the year 2030, the United Nations described data as the “lifeblood of decision-making and the raw material for accountability” for the purpose of realizing these developmental goals<sup>2</sup>. This curious yet key link between these new developmental goals and the use of quantitative data for agenda setting invited a flurry of big data-led initiatives such as but not limited to Data2X<sup>3</sup>, that sought to further strengthen and solidify the relationship between ‘Big Development’ and ‘Big Data.’

One of those SDG goals (Goal 5) prioritizes gender equality and empowerment of women and girls not only as a standalone goal but also as a crucial factor to realizing the other goals<sup>4</sup>. In response, several academic and non-profit initiatives have begun to interpret and conduct data-led gendered development or the “gender data revolution”. As with other data discourses, the gender-data discourse is also one of ‘speed’, charging ahead using a variety of quantitative and visualization approaches to *reveal* and eventually *solve* gendered problems of development. These interventions also invite some classical critical questions: *who is setting the agenda for the gender data revolution and who are its imagined subjects? How are questions of participation and asymmetries of power in developmental research being addressed? How does the gender data revolution address the situatedness as well as incompleteness of data records in the Global South (where most sites of intervention are)?* Speaking specifically to the theme of this special issue (‘cross-cultural feminist technologies’), this paper demonstrates how the welfarist discourse of data-led gender development is, in fact, assembled through the overwhelming enumeration of female-identifying bodies in the Global South. The paper offers critical historical insights from the fields of international development, anthropology, and postcolonial history to caution against both, the possible harms of gender disaggregated datafication as well as the consequences of non-participatory datafication of women, the subjects of the gender data revolution.

In the first part, this paper surveys some of the initiatives that began to interpret and materialize what data-led gendered development, also known as the ‘gender data revolution’ could mean and how it might be achieved. Then, the paper offers critical reflections on specific gender-data projects but also on the larger trend of portraying gender and sex-disaggregated quantitative research and intervention as a silver bullet for gender(ed) empowerment. Finally, the paper reflects on the uneven geographies of datafied development in the hope to course correct the onward march of the ‘gender data revolution’. Two clarifications are in order first. I interchangeably refer to the paradigm of data for gender(ed) development as ‘gender data revolution’ in the rest of the paper. I also use the term ‘gender data assemblage’ towards the latter half of the paper. Simply, I extend Anke Schwittay’s notion of the ‘financial inclusion assemblage’[1] that also draws on anthropologist Tania Li’s conceptualization of the assemblage

---

<sup>2</sup> United Nations Independent Expert Advisory Group. (n.d.). *A World that Counts: mobilizing the data revolution for sustainable development*.

<sup>3</sup> Data2X, per their website, is “an independent technical and advisory platform” housed at the United Nations Foundation supported by the William and Flora Hewlett Foundation and the Bill and Melinda Gates Foundation.” It was first announced in 2012 as an initiative by then US Secretary of State Hillary Rodham Clinton.

<sup>4</sup> Women, U. N. "Sdg 5: achieve gender equality and empower all women and girls." *New York: UN Women* (2017).

as “a set of practices that encompass things, subjects and organizations as well as various systems of knowledge, objectives and regulations. [1, 2]” The ‘gender data assemblage’ in this case refers to the coalescence of various actors in international development and global knowledge economies, local and national governments, activists, women and their families – that are all arranged in specific relationships so as to render social problems (“women and girls’ empowerment”) amenable to *specific* technical solutions (generally through datafication and specifically through gender or sex-disaggregated data).

In recent times, the proliferation of Artificial Intelligence (AI) and big data technologies in public life has invited substantial scrutiny, criticism, resistance and even refusal of technologies such as facial/affect recognition. Intermittently, researchers and activists have also warned about the disparate impact of blockchain based surveillance<sup>5</sup> as well as dataveillance on vulnerable populations such as gender minorities, non-white people as well as migrants and refugees. Even so, somehow, critical learnings on group data privacy<sup>6</sup>, contextual data privacy, ethical data collection practices, ethics of participatory research (“Nothing About Us Without Us”)<sup>7</sup> as well as recent learnings on computational bias seem to *not* carry over to developmental research and interventions. To some extent, this gap is also evidence of what constitutes “edge cases” or limits of disciplinary and institutional concerns whereby calls for data privacy, bodily integrity, participatory change and the assertion of rights-based frameworks find more purchase in Global North contexts while large scale data experiments through State-corporate-academia nexus remain unquestioned if they make developmental claims in global South contexts.

It is hard to interrogate and call for accountability in data-for-development initiatives as they not only span across national jurisdictions but also rely on specific directional flows of capital, power and decision making whereby they are at the same time treated as serious change efforts with abundant funding but the knowledge they produce is also not beholden to any specific organization, government or civic body.

With such awareness, this paper offers a critique of the contemporary ‘gender data revolution’ as well as the solutionist thrust of ‘gender disaggregated data’ within development. The former equates enumeration with inclusion and the latter argues that without gender-specific enumeration and hypervisibilization, quantitative data cannot be fully harnessed to empower women and girls in developing economies. I begin by tracing the discourse of ‘data in development’ and past calls for ‘gender mainstreaming’ in development to locate how two generally vast objects (gender and data) are brought into very specific relationships to attain the goals of development (i.e. ‘gender disaggregated data for development’). I then review case studies from Data2X’s 2017 report titled ‘Big Data and the Well-Being of Women and Girls’<sup>8</sup> to raise some questions about gender, data sources, methodology and goals, followed by broader critical reflections on similar initiatives. Finally, the paper offers a discussion on ‘big data solipsism in

---

<sup>5</sup> Pisa, Michael. "Reassessing expectations for blockchain and development." *Innovations: Technology, Governance, Globalization* 12.1-2 (2018): 80-88.

<sup>6</sup> Taylor, Linnet, Luciano Floridi, and Bart Van der Sloot, eds. *Group privacy: New challenges of data technologies*. Vol. 126. Springer, 2016.

<sup>7</sup> "Nothing About Us Without Us!" (Latin: "Nihil de nobis, sine nobis") is a slogan used to communicate the idea that no policy should be decided by any representative without the full and direct participation of members of the group(s) affected by that policy. The term became further popularized by its use in disability studies and activism in the speeches of South African activists Michael Masutha and William Rowland. It was then written about by James Charlton in his book ‘*Nothing about us without us: Disability oppression and empowerment*.’ Univ of California Press, 2000.

<sup>8</sup> Vaitla, Bapu, et al. "Big data and the well-being of women and girls: Applications on the social scientific frontier." (2017).

welfare' and the role of the 'gender data revolution' in rehearsing and reifying multi-layered geographies of uneven development.

## 2. Background: Sustainable Development Goals and Governance by Numbers

In 2015, at the United Nations General Assembly, 193 governments committed to a 15-year global development framework called the Sustainable Development Goals (SDGs). Also known as the 2030 Agenda or Global Goals, the SDGs will inform significant policy and programmatic interventions and resources through 2030[3]. Sustainable Development Goals (SDGs) replaced the eight Millennium Development Goals (MDGs) (Millennium Development Goals) that were set up by the United Nations as ambitious targets to be achieved by 2015. In 2015, while the MDGs were far from accomplished, in part because of how ambitious and general they were (for e.g. one goal was to eradicate extreme poverty and hunger), the Rio+20 conference (the United Nations Conference on Sustainable Development) in Rio de Janeiro, June 2012, spurred a process to develop a new set of Sustainable Development Goals (SDGs) to set up global developmental targets for 2030. They aim to “carry on the momentum generated by the MDGs and fit into a global development framework beyond 2015 [4].” From 8 MDGs, the number of Sustainable Development Goals increased to 17, retaining “gender equality and empowerment of women” as a top independent goal to achieve by 2030. Since their announcement, the SDGs have been critiqued[5] by some who have called them an “aspirational high school wish list”[6] due to the exhaustive detailing of goals. Other critiques of SDGs highlight the lack of a realistic plan to achieve the goals and, how human rights concerns become secondary to economic concerns within the imagination of global development[7]. There has been some criticism of the catchphrase that drove the vision of SDGs (“leave no one behind”)[8]. Fukuda-Parr and collaborators [9-11] who have written about MDGs in the past, lauded the SDGs for two reasons compared to their predecessors. Firstly, as they explain, the MDGs focused on ending absolute poverty as the primary goal, shifting the onus of action on Low Income countries, creating a North-South divide but also glossing over the role and concerns of Middle-Income countries. In contrast, SDGs go beyond economic prosperity to equally prioritize SRHR (sexual and reproductive health rights), access to justice among other outcomes. Secondly, as Fukuda-Parr notes, while “the MDGs were driven by leaders of bilateral and multilateral development agencies, the SDGs were mandated by Rio+20, from the UNCED process with broader constituencies...[12]” involving much broader diverse participation from activists, academics and civil society groups. However, in the same article, the authors also ponder the role of quantified measurement and how measuring certain SDG indicators *and* the foregrounding of measurable knowledge might shape the realization of transformative justice that is purportedly the end goal. As they say, while it is generally understood that numbers can help communicate a development agenda better, infusing it with a certain “scientific certitude”, “...in reality, quantification reduces complex and intangible visions – such as development that is inclusive – into concrete measurable objectives such as all children in school. [12]” As we also know from scholarship across disciplines (from STS to

sociology of knowledge to philosophy of science and more)[13-16], quantification of phenomena does not simply produce a different kind of representation but also creates a hierarchy of knowledge as well as limits of what can be quantified, and hence represented and what gets left out of the narrative. Historically, in global development as well, quantification of goals as well as the quantitative measurement by metrics has been reported to lead towards an overemphasis on target achievement and a general shift towards target-driven strategies rather than attention to the actual issues they set out to solve[14, 17]. While it is too early to measure anything (the SDGs came into effect in 2015), researchers following SDG implementation have asked how measurement and knowledge politics may affect the implementation of the ambitious and expansive SDGs with 17 goals and 169 targets and, in fact, whether “the power of numbers” will be implemented in selecting, simplifying, prioritizing and designing initiatives.

The SDG era also marks a crucial methodological shift where while previous development agendas were qualitative statements of social and political priorities, with the SDGs there has been a shift towards goal setting as a way to generate norms and overall ‘governance by numbers’ [12]. While quantitative targets were often included in past agendas, only a few were marked as actionable priorities but the SDGs mark a more systematic shift towards the adoption of the “language of numbers to articulate global norms [9, 18].” While the use of numbers as a tool of governance is a widely studied topic[19-21], the “proliferation of benchmarking and performance ranking across countries”, as Fukuda-Parr et al note[9], has led to growing research on governance by numbers – a practice that is itself gaining predominance in global governance. These studies[13, 14, 22, 23] show that the use of quantitative indicators is a tool of governance that has unique properties, exerts influence in particular ways, leads to distinct effects, and engages a particular type of politics. They warn of effects that are unintended or hidden, that are not part of the stated policy objective. Indicators are seemingly neutral but have deep effects on re-conceptualizing norms and shaping behavior that are not always visible, articulated, or benign.

My intent with summarizing some of the scholarship on SDGs is to establish how a fair deal of thought and anxiety already accompanies the discourse of enumeration within current global development efforts. At first glance, it is hard to trace the emergence or discursive formations that allow for data-based interventions or enterprise, including the ones that focus on gender and development. Fukuda-Parr and McNeill end their discussion with multiple observations, two of which are particularly relevant to this paper going forward. While they seem generally supportive of big data contributions, they call for the UN to play a proactive role in governing the use of big data to ensure inclusive and participatory development. They also highlight how at a time when overall funding for public statistics is drying up, organizations such as the Institute for Health Metrics and Evaluation (IHME), funded substantially by the Bill and Melinda Gates Foundation, have been on the rise in determining and advancing a specific focus within SDG indicators. There is then a larger discussion due, not only about the unevenness of North-South relationships in development but also the direct and indirect role of international philanthropic entities in generating global governance agendas. These institutional nodes are key actors in determining the *what* and *how* of global knowledge and studies of datafication and development must attend to the extension of institutional hegemonies through so-called “neutral numbers” that are in fact generated, disseminated and acted upon only through the auspices of global structural inequalities. I return to this issue in the latter half of the paper.

### 3. Gendered Development and the role of data

The specific SDG that this paper focuses on is the standalone SDG (Goal 5) that upholds gender equality and the empowerment of women and girls as critical to furthering global economic progress and human development. To be clear, the only explicit articulation of technological goals within its indicators is the one that calls for increased mobile phone access among girls and women<sup>9</sup>. Since one of the goals of this paper is also to produce a historical picture of concerns around enumeration in gendered development, this section provides a brief and selective summary of past efforts at “gender mainstreaming” within developmental agendas. What remains constant through sections on data in global governance, gender mainstreaming as well as the big data in gender development efforts, are implicit and explicit suggestions that there simply isn’t *enough data*<sup>10</sup> and without that development goals cannot be fully accomplished.

There is a good deal of literature on the push for ‘gender mainstreaming’[24] that first gained traction at the U.N. Decade for Women conference in Mexico in 1975. Interestingly, as Hillary Charlesworth notes, “mainstreaming” as a term itself was imported from the educational sector where it was used to make a case for integrating students with special learning needs into the mainstream classroom[24].” In the UN context, ‘gender mainstreaming’ came as a part of a larger paradigmatic change from ‘Women in Development’ (WID) to ‘Gender and Development’ (GAD) with the acknowledgement to focus on the relations between men and women in development policies. Gender mainstreaming was eventually adopted by the UN ECOSOC<sup>11</sup> that reaffirmed gender equity as the goal of gender-mainstreaming. The mainstreaming turn, of course, was still at a time when there wasn’t adequate visibility and understanding of the special burdens that women face with respect to developmental challenges. However, as Charlesworth noted, “...attention to questions of women and gender in the U.N. human rights system has been haphazard[25]” explaining that while special attention to women’s position in particular contexts did increase, it was mainly in statistical terms, which in the absence of contextual knowledge of gender roles and stereotypes, was inadequate to actually *affect* the human right in question. Two observations by Charlesworth remain pertinent to the new data-for-gender discussions. Firstly, the sex/gender distinction was elided in gender mainstreaming. Both, older literature as well as the new reports on data and gender interchangeably use sex and gender to talk about the value of disaggregation. Secondly, in practice, mainstreaming also created a certain kind of instrumentalization of the gender-data-development relationship where, for instance, as Elizabeth Harrison demonstrated through her germinal work[24, 26, 27], the “gender policy” pushed top-down by the Food and Agricultural Organization (FAO) based in Rome not only appeared irrelevant and inconvenient to local stakeholders in a fish farming project in

---

<sup>9</sup> List of indicators and targets under SDG5: <https://medium.com/sdgs-resources/sdg-5-indicators-5fe7d2b13b58>. Only indicator 5.b.1 explicitly engages with digital technologies (giving mobile phones to women)

<sup>10</sup> For instance, this report by the policy thinktank Brookings:

<https://www.brookings.edu/blog/techtank/2019/11/20/bridging-the-gender-data-gap/>, this report by the World Economic Forum that relates the digital and financial divide and attributes it to the lack of gender data:

<https://www.weforum.org/agenda/2018/08/we-cant-tell-if-were-closing-the-digital-divide-without-more-data/> and yet another by a UK based policy outfit Agulhas, connecting data gaps to gender development

<https://agulhas.co.uk/data-gaps-gender-development-reflection-gender-mainstreaming-international-womens-day-trisha-chuahan/>

<sup>11</sup> ECOSOC: (Economic and Social Council)

sub-Saharan Africa but also, the sudden number-collection on women threatened to disrupt the relationships between local male and female project workers. One final critique that I want to flag from the gender mainstreaming moment is that of Baden and Goetz, who remind us that the document resulting from the historic Beijing Conference of the World Bank, where gender mainstreaming was ratified as a goal, made the case for gender “...almost entirely on efficiency grounds, constructing a convergence between the interests of women and the promotion of economic liberalization...(World Bank, 1995:5)[28]” The point to remember here is not whether gender mainstreaming was a success or failure but rather what terms and promises were implied while historically advocating for an exclusive focus on women as developmental subjects, where such mandates are designed versus where they are implemented and, what exactly is imagined as successful gender development<sup>12</sup>.

Jumping forward to the SDG moment (2015), the UNDP (United Nations Development Programme) report on Gender and Development states,

*“In September 2015, Member States of the United Nations adopted, by consensus, the 2030 Agenda for Sustainable Development...The data revolution was recognized as an enabler of the 2030 Agenda. It can not only help to monitor progress towards the SDGs, but it also engages multiple stakeholders to advance evidence-based policies and programmes aimed to reach the most vulnerable and leave no one behind.”* [29]

Similar to Fukuda-Parr et al’s observations, especially in the domain of gender and development, the term ‘data revolution’ has come to gain currency but it should also be seen in continuation with the concerns raised earlier about the trend of ‘governance by numbers’ within developmental goal setting. As I show in the next section, the ‘data revolution’ *qua* ‘gender data revolution’ has begun to materialize from a broad, vague social aspiration into specific kinds of techno-social experiments.

## 4. Big Data for Gender Development: the new ‘Gender Data Revolution’

Since the emergence of Data2X and other smaller entities, we have begun to get a flavor of what gender or sex-disaggregated data might tell us about “women’s plight” through a series of published case studies and small projects. Given the scope of this paper, I discuss a few examples from one set of case studies that are illustrative of the phenomenon and its potential impact on women’s empowerment<sup>13</sup>. In this section I discuss case studies from the ‘Big Data and Well-Being of Girls’ report published by Data2X (2017)[30] that are in the public domain. To reiterate, I treat these projects as exemplars of what is a growing discourse and the purpose of my critique is not

---

<sup>12</sup> Schwittay’s paper on the ‘financial inclusion assemblage’ delves into the construction of the ‘poor brown woman’ as an ideal financial inclusion and development subject in detail. See Schwittay, 2011.

<sup>13</sup> At the time of writing this paper, only the first report i.e. ‘*Big Data and the Well-Being of Girls*’ (April 2017) published by Data2X was available to understand how the ‘Gender Data Revolution’ might look in practical application. Since the writing of this paper, Data2X has published a new set of case studies titled ‘Big Data, Big Impact? Towards Gender-Sensitive Data Systems’ (November 2019). While the author has not reviewed the new case studies in detail and they are beyond the scope of this paper, it is worth noting that methodologically, the new studies are similar to the ones discussed here (in their use of geospatial data, mobile data, social media data).



only to problematize these exact projects but rather to ponder aloud the scope and effects of the ‘gender data assemblage’ being made normative through them.

Within the SDG agenda, the lack of complete and reliable data on women globally has been recognized as a major impediment to designing informed policies for women’s progress<sup>14</sup>. As a blogpost by the UN Data Revolution team suggests, for the Data Revolution to become a ‘gender data revolution’... “*Women need to become more visible in data...*” and “*...women need to become active users of data.*” The proposition flows thus: *That in order to attain the goal of empowering women and girls as a part of the Sustainable Development Goals, there needs to be better and reliable data available on women (among other things).*

The data revolution then, is premised upon the “explosions in the volume of data, and the range of issues on which there is data. It seeks to capitalize on the volume of data, new data sources such as citizen-generated data as well as new producers of data (including those who use social media, credit and debit cards, smartphones and “smart” hardware). Especially, to usher a “gendered” data revolution, the UN report suggests that big data analytics can facilitate real-time situational awareness to “shine a light on the invisible” (by giving more information on the lives of girls and women), identification of regional trends, opportunities for real-time feedback and participatory monitoring among other things. The rationale, as the report explains, is that traditional data sources (such as household surveys, institutional records or censuses) are collected with a specific intention and they follow a structured format. As opposed to this, big data sources offer real-time, granular data on individuals as well as communities that can be analyzed to find a range of patterns post-collection. In that sense, it is suggested that the possibilities of what we can learn from big data streams (such as social media feeds from a region, call records etc.) by triangulating it with SES indicators as well as the gender of users, are endless, especially for an area or an indicator where traditional data sources don’t exist.

Data2X, a pioneering organization in this space, has identified what ideal gender data should look like or what kinds of gender data should be available. Building on the data needs identified by the WHO and ILO, a Data2X report titled ‘Mapping Gender Gaps’[31] states that these five desirable features – quality, coverage, standards, complexity and granularity, can help us identify and qualify types of gender data gaps. The same report identifies gender data gaps across the domains of global gender data gaps in health, education, economic opportunities, political participation and human security. For context, some of the main actors framing and pushing the ‘gender data revolution’ agenda include the UN Data Revolution team, Data2X (with support from the William and Flora Hewlett Foundation and the Bill & Melinda Gates Foundation) and the United Nations Global Pulse – a data and AI first innovation initiative that also work on gender issues. There are also smaller organizations such as Flowminder.org, regional non-profits, individual academics and university partners mostly located in the global North working on specific gender and data challenges. I begin with a Data2X report titled ‘Big Data and the Well-being of Women and Girls: Application on the Social Science Frontier’ where the authors summarize four projects that use various big data sources to “shed visibility” on women’s developmental indicators globally. Thematically, focusing on gender (or sex) disaggregated data – data on individuals broken down by gender or sex, the same report emphasizes how gender disaggregated data is key to realizing

---

<sup>14</sup> <http://www.undatarevolution.org/2014/12/15/gender-data-revolution/>

the larger mission of filling traditional data gaps on women and, in turn, empowering women in data-poor countries by informing regional policy-making.

Overall, the Data2X report identifies three main sources of big data to fill in the gaps within existing household survey data and national and institutional records, especially where there is no gender disaggregated data: 1) geospatial data 2) digital “exhaust” such as credit card activity and cell phone records and 3) public internet activity (mostly twitter data). In each category, there are examples of how disaggregating source data (geospatial, credit card and cellphone records as well as social media feeds) can tell us more about women’s challenges, their access to health, public transport etc. as well as regional-level trends on women and girls. Here, I briefly discuss the projects by data type and then raise overall sampling and methodological concerns observed across projects to anticipate the consequences of governing with big data.

## 5. Case Studies from the ‘Gender Data Revolution’ – using geospatial, digital exhaust and social media data

The first category is Geospatial Data where Flowminder researchers use satellite imagery along with DHS<sup>15</sup> household survey data in order to improve the spatial resolution of existing data on girls’ stunting, women’s literacy, and access to modern contraception in Bangladesh, Haiti, Kenya, Nigeria, and Tanzania. As the report describes, the project leverages the fact that many types of social and health data are correlated with geospatial phenomena. In this project, researchers developed modeling techniques using publicly available high-resolution geospatial data to then infer “similarly high-resolution patterns of social and health phenomena across entire countries.” Based on their results, such mapping seems to reveal more about the scope of modeling than actual correlations per se. For instance, the authors note that model performance varied greatly across indicators and countries. Especially for girls’ stunting, models proved inadequate in all countries but Nigeria. Similarly, when trying to map modern contraceptive use through geospatial data, models performed strongly in Tanzania and Nigeria, mainly because, as the authors note, the set of optimum geospatial variables depends on context. The project’s overall findings were able to extrapolate available survey data to various indicators (accessibility, literacy etc.) to produce larger, region and nation-wide maps of three indicators in Nigeria and literacy in Kenya. While a detailed analysis is beyond the scope of this report, what remains relevant to this paper is the fact that models of various women’s variables could not be produced in areas where there wasn’t already existing indicator data (for instance, contraceptive use in Bangladesh and Kenya). Further, another issue the authors note is that some of the models they attempted were inaccurate in their prediction of well-being outcomes and would need more survey data to improve model performance. Finally, and critically, they remark that, “the exact nature of the relationships between geospatial phenomena and well-being outcomes is unclear.”

---

<sup>15</sup> Domestic Household Survey or DHS is a USAID funded program that has been providing technical assistance and aiding the collection of health and population trend data through largescale surveys in over 90 countries.

A few remarks are in order. Responding to both, the (big) data gaps as well as speculating as to why data on certain indicators might be missing for some countries, we wonder if engagement with historical as well as qualitative accounts of women's health, cultural and religious perceptions and contextualized gendered living might not have helped in filling these gaps. It is also not explained why this specific exercise did not factor in religious distribution among the variables it considered, which might have been pertinent to the mapping. Finally, not to dismiss the value of wide scale mapping, we want to echo the researchers' own claim of "granular mapping" to ask how big data analysis might attend to smaller socio-economic and political forms such as the 'community'. Even so, if this experimental and limited project was presented to a local or regional government agency, there is a good chance that local authorities are broadly aware of the attitudes towards contraception, female health, literacy and so on as well as the lack of, or disparate distribution, of resources.

There is also a different kind of causal implication or a qualitative concern to the kinds of social interventions that might be based on figuring out where and why variables don't correlate. For instance, when the report notes, in the case of girls' stunting and correlation with economic inequality, "the areas with higher absolute levels of girls' stunting...are not necessarily the areas of greatest inequality; the northeast, central, and southern urban regions of Nigeria appear to exhibit the largest disadvantage for girls (ibid)", to which the next step would be to ask *why* these don't correlate and in-turn, focusing in on regions that stand as outliers. That is a very specific kind of inquiry and necessarily produces an economy of attention, intervention and aid depending what one concludes from a correlation. Of course, as the authors agree, the same two variables could have numerous relationships in different regions. Reflecting on the larger goal of gender-disaggregated data and women's empowerment, while geospatial mapping could produce a tentative yet expansive view of various welfare indicators, I want to highlight the foregrounding of contextual interventions as well as the need to discuss how such models can interact, leverage and support other methods (qualitative and quantitative) in order to produce not just accurate but also meaningful correlations, a concern that I address at length later in the paper through scholarship in feminist geography. An important goal of the Data Revolution is also to "put the data in the people's hands" so to speak. This is a foundational concern that needs to be addressed while building geospatial models – by involving communities, families and administrative agencies – all of whom will be crucial to providing a relational context for variables, revealing unmapped factors in the DHS data and crucially, in what kind of information they really need and lack in order to support girls and women's welfare.

Moving to the next category, 'digital exhaust', the Data2X report explains,

*"Digital technologies are ubiquitous, and their use leaves traces—records of the goods and services we consume, the places we go, and the people with whom we interact."*

This trace information is broadly defined as 'data exhaust' or 'digital exhaust' and includes cell phone records and credit card activity as two vital data sources. The report describes a project undertaken by the (Data2X) researchers themselves, where anonymized individual CCR (credit card records) and CDR (call detail records) activity collected over a 10-week period, from 150,000 users, with associated age, sex and location information was analyzed to examine expenditure priorities and mobility patterns across different age groups, income levels and sexes. The authors clarify that the CCRs "include data on the broad types of goods and services purchased,

expenditure amounts, and the chronological sequence of transactions.” Only for 10% of the credit card users in the data source also *had* call detail records. Some of the project’s findings present differences between men and women commuters’ mobility levels, also stating that men tend to live closer to city centers and have better access to economic opportunities. They observe that women homemakers tend to be less social (where social means outgoing) and have reduced mobility compared to men. Same holds for young women compared to men.

While in principle, the report’s findings are insightful, an obvious question that is about the usefulness of sampling mobile phone records, or *whom* one intends to know about by querying a certain big data source. Things that should have been explained about the project include the national demographic context, which layer of the population the project sought to investigate, why and what the limitations of the data source and analysis might be. Two immediate issues are apparent if one seeks to produce knowledge about women especially in purportedly uninformed geographies: 1) the utility and reliability of cell phone and credit card activity as representative sources and 2) the need for a thorough explanation of the institutional and infrastructural partnerships as well as consent-taking mechanisms in using what is essentially also surveillance data. The project and the report do not address these issues, nor do they reflect on the consequences of hyper-visibility that such mapping may create in its wake.

The third and final category is ‘social media (data)’. In two different projects (one from the Data2X report) that I use as examples, researchers utilize publicly available social media data (twitter and Reddit posts) in English language to map mental health as well as public perception on various issues. The first project (not in the Data2X report) [32] analyzes 1.5 million posts from four countries using machine learning techniques, to “identify genuine self-disclosure of mental illness from social media posts.” The project’s rationale is that overlooking sex-based differences in mental health can have drastic consequences and prohibit help-seeking. They also compare modes of linguistic expression and topical content along gender/sex to reveal differences in sex- and culture-based expressions of mental health concerns. The second project conducted by Data2X researchers in collaboration with the UN Global Pulse team and the University of Leiden sought to develop and prototype a tool “to infer the sex” of social media users using publicly available Twitter data [33]. Using machine learning techniques, the tool can predict the gender of a user account with a precision of 74-87%. It uses open source code from an earlier tool called the ‘Gender Computer’ developed by TU Eindhoven researchers<sup>16</sup>. From their own description, the gender computer is a tool that “...tries to infer a person’s gender from their name (mostly first name) and location (country)” The end goal of both the old and new tools is to help associate gender in real-time to provide a gendered context to social media expression.

Again, some fundamental concerns emerge. The first study, or at least its results section devotes no attention to discussing the implications of studying cultural and linguistic expression from four different countries. In terms of limitations, they acknowledge that the terms used to filter mental health concerns are not exhaustive. Further, they also recognize that their sample is not representative of the general population. In fact, as they say, “...Twitter users (which) are likely to be more affluent, more technologically skilled and more willing to express themselves about mental health issues...” However, they make no mention of the implications of studying social

---

<sup>16</sup> The source code for the gender classifier used in these projects can be found here: [https://github.com/LU-C4i/gender\\_classifier](https://github.com/LU-C4i/gender_classifier) and the ‘Gender Computer’ tool can be found on Github: <https://github.com/tue-mdse/genderComputer>

media data as representative of such diverse socio-economic populations<sup>17</sup>. Further, while the authors talk about men, women, youth and shuttle between developing and developed countries (India, South Africa, UK and USA), the illustrative images that accompany their findings show a woman in some kind of a non-Western geography with a low-cost feature phone[34], who by no means invokes the imagination of a Twitter or Reddit user, even in a developing country. Throughout the report, more such images of women and children in developing world settings are deployed alongside the findings, invoking developmental tropes of poverty and lack that have already been sufficiently critiqued in the past. It is understandable that all the projects I reviewed here primarily serve to project the potential of data for gendered development. However, such misaligned images as well as various disclaimers in each project keep shifting the target of what big data can do and what it actually does through each of the projects discussed here. There are also more general concerns around sampling and methods in all projects, such as how representative any of these datasets are of the respective developing country populations and even further, of women in these countries, already recognized as a minoritized and often “uncounted” group. If the original claim of the “data revolution” agenda is to capture and render visible the reality of girls and women in traditionally unmapped areas then it is not enough to simply state the limitations of data sources (CDRs, CCRs, social media...). In fact, the more important discussion to be had is – what kinds and combinations of data sources and methods might most comprehensively cater to the questions that the gendered data revolution claims to answer<sup>18</sup>. If women in developing countries are already minoritized in terms of their participation in the formal sector, access to financial services and technology, then it is not a limitation but should rather be the guiding principle of D4D project design. As to the final project that “computes” or disaggregates the gender of social media accounts, there is a problem for those whose names defy the (Western and binary) computational logics of the gender computer. Even for those whose names can reliably be used to assume their (binary) gender, the essentialist logic of biology or static culture-based behavioral differences (men say X and women say Y) is incredibly problematic at a time when binary gender assumptions are under question.

## 6. Broader Critical Reflection

In an article titled “Sex disaggregation will not energize equality” Sarah Bradshaw[31] offers a compelling critique of gender disaggregation efforts within energy research and development where she cautions against the flippant interchangeable use of gender and sex as categories within mainstreaming discourse. These projects seem to be “doing gender” in the sense of highlighting the activity and traits of girls and women, implying that there is a difference between what men and women do (although men’s data is never explicitly discussed in the projects I reviewed). Echoing what Bradshaw notes, merely visibilizing difference means little without contextualizing gendered differences and gender roles and relations as well as the reach of

---

<sup>17</sup> Eszter Hargittai discusses the potential and challenges of social media as a data source in detail. See: Hargittai, Eszter. "Potential biases in big data: Omitted voices on social media." *Social Science Computer Review* (2018): 0894439318788322.

<sup>18</sup> Also see Patton et al’s paper on annotating social media data from vulnerable populations with the help of graduate students versus domain experts/activists: Patton, Desmond, et al. "Annotating social media data from vulnerable populations: Evaluating disagreement between domain experts and graduate student annotators." *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 2019.

institutions like families, communities and governments in maintaining, causing or disrupting those gendered differences. As she says, “Explaining differences by gender thus needs an understanding not just of gender theory, but of the specific context in which the gender differences occur. (ibid)” Such ventures get further complicated when we locate the power differentials between those studying and those being studied. As Bradshaw says, “It is strange that a western man, or woman, can think they ‘know’ through their own experiences about, for instance, rural women’s experiences of inequalities in the Global South.” Echoing her concerns, it is worth asking the same of the project discussed above, as to why an intersectional view of gender wasn’t taken across these initiatives. It is also unclear why, despite there being an abundance of global intersectional feminist scholarship, the nuances of intersectional as well as situated gender research do not trickle through in development initiatives such as the ones discussed above. Not only do women not live in isolation but also their gendered oppression and empowerment are necessarily connected to family, communities, religion, race and the socio-political and economic milieus they inhabit. But importantly, the argument here is not for simple inclusion – the point is a more pedagogical one where I am asking how an expansive intersectional understanding of complex social realities and specificities might complicate and offer new challenges for gender disaggregation among the global North researchers leading the pack on the issue. For future ethnographic research, it would be valuable to understand how researchers engaged in such interventions imagine and account for the harm and risk that datafied visibilities may bring to the local contexts in which they intervene through the “god view” of data dashboards.

## 7. Discussion

Beyond the case studies discussed above, I offer two points for discussion here. While this paper does not deal with datafication within the context of state boundaries, the intent behind tracing the role of data in gendered development through international goal setting is to locate the global power geographies that datafication reifies and/or disrupts. Especially, as reviewed early in the paper, if the SDG framework is considered more progressive and participatory in imagining development as a global responsibility and not as unidirectional North-South flows, then it is important to locate the sites, stakeholders and actors of such change. Developments in data infrastructure and data flows produce novel implications for how we imagine territorial sovereignty and how nation states govern their sovereign subjects. With the rise of cloud infrastructures and the subsequent political crises (and affordances) created and perceived by various governments, ruling parties and others, movements for data sovereignty and data localization (or “against data colonialism”) have gained momentum in different parts of the world. However, international data development projects such as the ones described in this paper occupy a distinct position that makes them somewhat immune to the scrutiny that typical interventions might attract. Of course, one reason is that it is hard to map the economic geography of international development along the lines of national boundaries. These philanthropic organizations as well as the UN agencies in each development sector may be geographically located in some places (often headquartered in the global North) but their claims and scope of interventions are global, sometimes even exclusive to global South territories in terms of action. But the flow of power, the centers of decision-making and the design of even

participatory development interventions have historically been North-South and top-down rather than horizontal or challengeable by the purported beneficiaries. The affordances of data forms and flows complicate these mappings even further because not only do data get produced, stored, shared, retrieved and acted upon in distributed ways but also, as all the projects that I discussed in this paper did, it is possible to act from a great distance with data. Historically, statistics and enumeration projects have come to be the “natural allies of colonialism” because of how numbers allow the powers that be to create abstract knowledge for colonial administrators to know *enough* to govern at a distance, without engaging in processes of consensus-making. In the context of international development, tomes have been written critiquing “helicoptering” interventions that operate from above, bringing in Western experts to measure, decode, frame and *formulate* the problems that they proceed to then solve. Against this backdrop, returning to the discussion on gender disaggregated data for (gendered) development, the two concerns I raise here are: 1) Big Data Solipsism in Welfare and 2) The Uneven Geographies of Datafied Development.

## 7.1. Big Data Solipsism in Welfare

In an article titled “Dangerous Data: the role of data collection in genocides” [35], researcher and activist Zara Rahman offers multiple examples of how data gathering on vulnerable groups especially in politically sensitive areas and situations has historically ushered danger and disaster upon the same communities at threat. She offers the examples of “ethnic ID cards” during the Rwandan genocide, comprehensive population registers appropriated to target Roma and Jewish communities during the Holocaust in the Netherlands and more recently, two separate exercises conducted by the Bangladeshi government as well as the UNHCR<sup>19</sup> to collect biometric identification data on the stateless Rohingya individuals in Bangladesh. Not just for identification but as colonial and postcolonial historians have shown, various technologies of self-making including fingerprinting, measuring intelligence as well as colonial census surveys have evolved into more sophisticated and intrusive technologies of contemporary dataveillance (such as the Aadhaar biometric project, forms of sentiment analysis and affect recognition tech as well as State-issued census). These technologies are pushed onto citizens and welfare subjects at large with the promise of development where development is defined as a more efficient and transparent distribution of welfare resources. The claim that we cannot empower girls and women if we do not know them fully – performs a similar function. Especially, in advancing the datafication and analysis of women’s plight through modes of knowledge that operate at a distance (tracking movement, mobile-use, social media activity etc.) with *no* “thick” or participatory engagement, the implicit claim goes even further. The suggestion then is also that the subaltern *cannot* speak – meaning that we cannot rely on developmental subjects to articulate their own needs and desires or perhaps, these subjects just cannot know enough or better than what the numbers and digital traces can tell us. To be clear, this is not an ungenerous analysis of the ‘gender data revolution’ since 1) data does not *have to* mean quantitative data and yet, 2) all the case studies illustrative of the ‘gender data revolution’ are *not* qualitative, not even mixed methods studies. Nor do they include community feedback or consultation after the studies were done.

---

<sup>19</sup> UNHCR: United Nations High Commissioner for Refugees

This paper tries to follow the creation of a 'gender data assemblage' whereby gender, quantitative data and powerful developmental institutions come to be situated in specific relationships of utility and interdependency that may eventually seem "natural" and self-evident. In response, Fuentes and Cookson note how the call for a 'gender data revolution' hinges upon powerful institutions claiming that "progress on gender equality depends upon our collective ability to close the 'gender data gap' by which they mean that we know less about women's lives than men's in statistical terms. [40]" Feminists within and outside the academy have called out such a "measurement obsession" [36] and have shown how such an approach is simply "ill-fitted to capture the complexities of gender-subordination.[37]" These criticisms are not new but also, over-reliance on "data truths" especially for the purposes of welfare has wrought havoc time and again. For instance, when the Indian government ruthlessly pushed for the enrolment of every citizen in its national biometric identification system (Aadhaar), approximately 8.8% of Aadhaar cardholders reported errors in their IDs<sup>20</sup>, which given the population of India, amounts to 120 million people who risked losing access to bank accounts, government subsidies and more. In another financial inclusion scheme, called the 'Pradhan Mantri Jan Dhan Yojana' (PMJDY) under which the number of Indian adults with a bank account went up from 53% in 2014 to 80% in 2017[38], based on banking data, 48% of those who made an account in a financial institution had made no withdrawal or deposit between 2017-18. Based on the known endemic factors that hinder access to formal credit especially among minoritized sections of the country, critics of the scheme had already forewarned of this possibility – of the eventuality that merely an increase in bank accounts may not translate to financial inclusion. In 2018, Rohingya refugees in Bangladesh even went on a strike demanding the government to stop further data collection citing fear that their data might be shared with the Myanmar government<sup>21</sup>.

Two important points emerge from this discussion. Firstly, the 'gender data revolution' behemoth furthers an unhinged death drive for datafication at its worst and ignorantly replicates the pitfalls of databased welfarism at its best. Explanations and clarifications for why this "revolution" disregards participatory engagement and ethnographic inquiry, are due. Secondly, in its wake, the gender data assemblage builds permanent information infrastructure that isolates, exclusively enumerates and depends on female-identifying bodies' lives. The resultant data and techniques can be leveraged to control and monitor women for any number of purposes in the future. Most importantly, since such a 'gender data revolution' is not responsive or responsible to concerns articulated by smaller, specific groups of girls and women, it can never fail! It exists in a state of permanent iteration, hopefulness and potentiality where the only solution to its shortcomings is "to get more data, to know more" while the responsibility for meting out the data revolution's promises returns to national and local actors who must negotiate and perform to align with *what the data tells us*. Anke Schwittay's paper on the 'financial inclusion assemblage'[39], discussed earlier in this paper is particularly instructive because, as she traces in her argument, a shift from the financial inclusion of poor (women) to the financialization of debt, in the 'gender data revolution' as well, the obsessive tendency for datafication has only reproduced women and girls

---

<sup>20</sup> Rajagopalan, S. (2019, December 9). Opinion: The inevitability of errors in determining citizenship. Retrieved January 31, 2020, from <https://www.livemint.com/opinion/columns/the-inevitability-of-errors-in-determining-citizenship-11575909663129.html>

<sup>21</sup> Bangladesh faces refugee anger over term 'Rohingya', data collection. Reuters. Mohammed Nurul Islam <https://www.reuters.com/article/us-myanmar-rohingya-bangladesh/bangladesh-faces-refugee-anger-over-term-rohingya-data-collecton-idUSKCN1NV1EN>



in terms of their utility for the development industry. It is not uncommon for developmental economists to talk about materially poor and other developmental subjects in the language of Social Darwinism<sup>22</sup>, as if, given the right resources and education, the materially poor will eventually evolve or emerge out of poverty. Such an approach discounts the role of structural historical oppression, cultural and political systems as well as the aspirations and desires of materially poor and minoritized communities of gender and color, assuming that every developmental subject must channel their lives to align with malaria prevention or banking or any other project imposed upon them. Without discarding the informational utility of data, the ‘gender data revolution’ could benefit from deeper and intimate engagement with social power in every case study to articulate ‘gender data gaps’ that align with gendered bodies and their communities’ concerns.

## 7.2. The Uneven Geographies of Datafied Development

Feminist geographers Fuentes and Cookson[1] offer a spatial critique of the ‘gender data revolution’ where one of the points they make is that the conversation on the gender data gap in development has been largely *aspatial* thus far, meaning that there is a lack of “quality statistics” and complete datasets to make global claims about women’s poverty, for instance. They draw on Bradshaw, Chant and Linneker[40] who point out how the claims that *are* made rest on incomplete regional datasets in which results may be skewed when some countries do not produce comparable data or when women are more likely to be poor in the majority but not all countries. They also raise the issue of imprecise units of measurement or the question of ‘data fit’ in claims such as “poverty is feminized” (implying that female-only or female headed households are particularly prone to poverty). As they rightly clarify, the consolidation of fuzzy, subjective and complex realities of household structures and the “diversity of the family form” to arrive at indicators such as ‘female-headed household’, collapses breaks or changes in family composition and also produces a kind of absolute – a classification and an exemplar subject for all future research and policy experimentation on so-called ‘feminist poverty’. They also point to how global and national indicators tend to “delegitimize and obscure non-dominant, ‘local’ forms of knowledge and understandings about the world”, a point raised in this paper as well. In short, the decontextualization of numerical representations and their abstract journeys from the local to the regional to national and then global agendas inheres the risk of statistics, already inadequate and dynamic, getting hijacked, co-opted and privileged over the contextual, political and social struggles that these numbers claim to represent. The question then remains is whether data experiments done in the name of the ‘gender data revolution’, drawing on incomplete data with multiple caveats (as I showed in some examples earlier), may serve any utility or meaning for local

---

<sup>22</sup> In 2020, developmental economist and Nobel Prize winner in Economics, Abhijit Banerjee who literally co-runs a “lab” for poverty eradication said in one of his public interviews that, “*If the poor are given assets, for example cows and goats, and basic training to start a business, over a period of time they inculcate a confidence to stand on their feet and make ends meet.*” His statement while perhaps well meaning is reflective of developmental language about the materially poor where the poor are talked about as a homogenous group with little understanding of their own “condition” (poverty) or the intellectual resources to solve/treat it. See his statement:

<https://www.outlookindia.com/website/story/india-news-i-would-be-petrified-if-says-nobel-laureate-abhijit-banerjee-on-citizenship-law/345182>

power struggles and challenges that non-male bodies face. Or worse, what if such metrics empower and institute yet another disciplinary mandate driven by faceless powerful “global” entities who tell women to start living to improve these metrics?

The other pertinent geographic/spatial point that Fuentes and Cookson offer is not just attention to the disconnect between numbers and contextual representation but also whether documentation (qualitative or quantitative) can be re-oriented, made creative, made political to start mapping socio-spatial practices and infrastructures rather than producing accounts of individual behavior. In particular, they offer the case of gender data and femicide in Guatemala to reflect on how, typically, data drives on gender-based violence focus on ‘prevalence data’ to account for femicide victims. The authors recount a case of Adela Chacon-Tax, a young mother of 3 and one of the victims who were “officially counted” in the statistical data on femicide. As Fuentes and Cookson [40, 41] argue, while Adela’s death was counted among prevalence data, the data did not capture the emotional and material labor of Rebecca, Adela’s sister who fought and waited for 3 year to get justice or the trauma of her eldest daughter who discovered her mother’s body in a ravine or the support by Rebecca in taking care of Adela’s children to ensure they did not become another vulnerable statistic. The point then is not to simply suggest that quantitative data collection is inadequate and futile but to ask what might happen if these realities could shape *what* data is collected and highlighted in the first place. What if each such case of violence shifts the focus from the female victim’s death or even interpersonal violence to the material and social infrastructures that enable gender-based violence (“expanding the catchment data”)? Such an approach pays attention to socio-spatial designs as crucial to understanding and ameliorating gendered lives [40].

Finally, one last structural point needs to be addressed in this critique, a structural problem that continues to undergird developmental enterprise despite the SDG’s claims of being a truly global agenda for development. As many have noted, there is a disturbing trend, particularly in the arena of agenda setting in international development in which expertise seems to be increasingly sought from corporate management consultancies such as McKinsey and Accenture. As they rightly state, “While these corporate actors are always ready to produce largely apolitical reports on, for example, the profitability of gender equality, they often lack training in feminist research and analysis...[42]” As such there is a vast body of critical work that addresses the problems of philanthrocapitalism and the immensely problematic roles of knowledge/power brokers such as but not limited to the Hewlett Foundation, Gates Foundation, Ford Foundation etc. Such international entities that are not directly bound to a single national mission or sovereign forms of accountability have been known to subvert developmental priorities through their funding and visibility mechanisms[40]. In this moment too, of the ‘gender data revolution’ and the resultant ‘gender data assemblage’ production, a cursory look at the Data2X 2019 report<sup>23</sup> (not same as the one discussed in depth in this paper) reveals that a majority of the grantees<sup>24</sup> are either based out of the Global North or are distributed across the world but headquartered in developed countries while a majority of the focus geographies of intervention are in the global South (Dominican Republic, Chile, Nigeria, Pakistan, Nepal, Uganda, India).

---

<sup>23</sup> <https://data2x.org/wp-content/uploads/2019/11/BigDataBigImpact-Report-WR.pdf>

<sup>24</sup> To name a few: TheGovLab/NYU, University of Bologna, University of California Berkeley, Flowminder Foundation, WorldPop, Oxford University, University of Southampton, Overseas Development Institute, Northwestern University, Women’s World Banking)

Not just in terms of representation or even participatory concerns, but something more serious is at stake here. Apart from the obviously skewed distribution of *who* gets to produce databased knowledge (Global North) and *where* data can be easily obtained, experimented and analysed without data privacy and security scrutiny, the case of the “gender computer” (discussed earlier in this paper) also raises concerns for intersectionality. The aim of the “gender computer” and the ‘gender classifier’ is to specifically arrive at male and female names and if need be, using facial recognition on profile photos to achieve more accuracy in these male and female classifications. It is not hard to see how such an enterprise might run into resistance and criticism especially when Global North cities are leading the charge on refusing intrusive facial and affect recognition technologies. It is also not a stretch to imagine that such efforts would have little purchase especially in Global North geographies where feminist and queer movements have made binary gender classifications antiquated and contestable, including in census exercises. However, the same cannot be said of global South spaces, not because they are in a state of cultural lag but due to multiple other factors (ongoing struggles for legal and social recognition of queer bodies, more urgent material priorities). Then, against such a backdrop, even the most intrusive and problematic data and AI solutions find purchase and are often tested on global South subjects in the name of development with limited or no accountability for consequences. The ‘gender data assemblage’ in its current form then replicates and reifies traditional ‘unevenness’ of development as it flows through familiar circuits of philanthrocapitalism and western academic and non-profit interventionism with little discussion of technological ethics or reflexivity extending to these ventures.

## 8. Conclusion

This paper builds on and recognizes the contributions by feminist scholars and activists on topics of gender mainstreaming and data-led solutionism in international development. The aim of this paper then, was not simply to repeat critical and historical critiques of solutionist development and the limits of enumeration at large but to also think through the challenges that computational logics and big data bring to the questions of gender(ed) development. If the case studies discussed in this paper are taken as indicative of the kind of knowledge and methodological approaches preferred by the institutions leading charge on the ‘gender data revolution’, serious concerns arise around the erasure of girls and women’s agencies in their own empowerment as well as the creation of powerful dataveillance pipelines that have been assembled on the backs of female-identifying data-bodies. Such an erasure and regime of perpetual potential surveillance unsurprisingly advances a patriarchal gaze and control over women’s futures. The new information and institutional infrastructure for such a ‘gender data assemblage’ is being assembled by a majority of Global North actors while its detrimental effects skew towards global South women and queer bodies who may not be able to afford protection and accountability against the long term effects of developmental programming and State imperatives shaped by the decontextualized gender data assemblage.

Calling for attention to “thick” and “small” data and demanding for equal attention to creative, feminist, decolonial and participatory methods in databased welfare might not appear to be an attractive alternative because it necessarily demands slowing down the speed of research and

interventions to first engage with the entanglements and frictions of lived experience and daily negotiations of power. To create a better world for women and their others then, the researcher *has to* account for their (developmental subjects') experiences and demands.

Erratum:

In the final publication published in April 2021, the author misattributed a quote on Page 16.

Further, after feedback from cited authors, the references on Page 18 have also been updated. We deeply apologize for these errors and thank the authors for their corrections.

## References

1. Schwittay, A.F., *The financial inclusion assemblage: Subjects, technics, rationalities*. Critique of Anthropology, 2011. **31**(4): p. 381-401.
2. Li, T.M., *The will to improve: Governmentality, development, and the practice of politics*. 2007: duke university Press.
3. Desa, U., *Transforming our world: The 2030 agenda for sustainable development*. 2016.
4. Vandemoortele, J., *Advancing the global development agenda post-2015: some thoughts, ideas and practical suggestions*. New York: UN System Task Team on the Post-2015 UN Development Agenda, 2012.
5. Waage, J., et al., *Governing the UN Sustainable Development Goals: interactions, infrastructures, and institutions*. The Lancet Global Health 3.5, 2015: p. e251-e252.
6. Hickel, J., *The problem with saving the world*, in *The Jacobin Magazine*. 2015.
7. Hawkes, S. and K. Buse, "Searching for the Right to Health in the Sustainable Development Agenda: Comment on" *Rights Language in the Sustainable Development Agenda: Has Right to Health Discourse and Norms Shaped Health Goals?* International journal of health policy and management 5.5, 2016: p. 337.
8. Stuart, E. and J. Woodroffe, *Leaving no-one behind: can the Sustainable Development Goals succeed where the Millennium Development Goals lacked?* Gender & Development 24.1 2016: p. 69-81.
9. Fisher, A. and S. Fukuda-Parr, *Introduction—Data, Knowledge, Politics and Localizing the SDGs*. Journal of Human Development and Capabilities, 2019. **20**(4): p. 375-385.
10. Fukuda-Parr, S., *From the Millennium Development Goals to the Sustainable Development Goals: shifts in purpose, concept, and politics of global goal setting for development*. Gender & Development, 2016. **24**(1): p. 43-52.
11. Fukuda-Parr, S. and T.S. Hegstad, "Leaving No One Behind" as a Site of Contestation and Reinterpretation. Journal of Globalization and Development, 2018. **9**(2).
12. Fukuda-Parr, S. and D. McNeill, *Knowledge and Politics in Setting and Measuring the SDG s: Introduction to Special Issue*. Global Policy, 2019. **10**: p. 5-15.
13. Broome, A. and J. Quirk, *Governing the world at a distance: the practice of global benchmarking*. Review of International Studies, 2015. **41**(5): p. 819-841.
14. Fukuda-Parr, S., A.E. Yamin, and J. Greenstein, *The power of numbers: a critical review of millennium development goal targets for human development and human rights*. Journal of Human Development and Capabilities, 2014. **15**(2-3): p. 105-117.
15. Hacking, I., *Making up people*. 1986.
16. Pine, K.H. and M. Liboiron. *The politics of measurement and action*. in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015.

17. Fukuda-Parr, S., *Global goals as a policy tool: intended and unintended consequences*. Journal of Human Development and Capabilities, 2014. **15**(2-3): p. 118-131.
18. Kanie, N. and F. Biermann, *Governing through goals: Sustainable development goals as governance innovation*. 2017: mit Press.
19. Poovey, M., *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. 1998: University of Chicago Press.
20. Porter, T.M., *Trust in numbers: The pursuit of objectivity in science and public life*. 1996: Princeton University Press.
21. Strathern, M., *Audit cultures*. Vol. 146. 2000: London: Routledge.
22. Merry, S.E. and J.M. Conley, *Measuring the world: Indicators, human rights, and global governance*. Current anthropology, 2011. **52**(S3): p. 000-000.
23. Merry, S.E., *The seductions of quantification: Measuring human rights, gender violence, and sex trafficking*. 2016: University of Chicago Press.
24. Charlesworth, H., *Not waving but drowning: Gender mainstreaming and human rights in the United Nations*. Harvard Human Rights Journal, 2005.
25. Charlesworth, H., *Not waving but drowning: Gender mainstreaming and human rights in the United Nations*. Harv. Hum Rts. J., 2005. **18**: p. 1.
26. Harrison, E., *Fish Feminists and the FAO: Translating 'Gender' through Different Institutions in the Development Process*. 1997: Zed Books.
27. Harrison, E., *Fish feminists and the FAO: translating gender through different institutions in the development process*, in *Getting Institutions Right for Women in Development. Reference, Information and Interdisciplinary Subjects Series*. 1997, Zed Books. p. 61-74.
28. Crewe, E. and E. Harrison, *Whose development*. An ethnography of aid, 1998: p. 23-65.
29. Baden, S. and A.M.G. Goetz, *Who needs [sex] when you can have [gender]*. Feminist visions of development: gender analysis and policy, 1998: p. 19-38.
30. U.N., *Transforming our world: The 2030 agenda for sustainable development*. 2015, Resolution adopted by the General Assembly.
31. Vaitla, B., et al., *Big data and the well-being of women and girls: applications on the social scientific frontier*. 2017, Data2X.
32. Buvinic, M., R. Furst-Nichols, and G. Koolwal, *Mapping gender data gaps*. Data2X report, United Nations Foundation, Washington, DC. [http://data2x.org/wp-content/uploads/2014/11/Data2X\\_MappingGenderDataGaps\\_FullReport.pdf](http://data2x.org/wp-content/uploads/2014/11/Data2X_MappingGenderDataGaps_FullReport.pdf), 2014.
33. De Choudhury, M., et al. *Gender and cross-cultural differences in social media disclosures of mental illness*. in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 2017.
34. Pulse, U.G. *Sex-Disaggregation of Social Media Posts*. Big Data Tools Series 2016 [cited 3.
35. Bradshaw, S., *Sex disaggregation alone will not energize equality*. Nature Energy, 2018. **3**(10): p. 813-815.
36. Rahman, Z., *Dangerous data: the role of data collection in genocides*. The Engine Room, 2016. **21**.
37. Liebowitz, D. and S. Zwingel. *Assessing Global Gender (In) Equality: A CEDAW-Based Approach*. in *annual meetings of the International Studies Association, Toronto*. 2014.
38. Buss, D., *Measurement imperatives and gender politics: an introduction*. Social Politics: International Studies in Gender, State & Society, 2015. **22**(3): p. 381-389.
39. Sanghera, T. *80% of Indians now have a bank account. So why is financial inclusion low?* Business Standard, 2018.
40. Fuentes, L. and T.P. Cookson, *Counting gender (in) equality? a feminist geographical critique of the 'gender data revolution'*. Gender, Place & Culture, 2019: p. 1-22.

41. Bradshaw, S., S. Chant, and B. Linneker, *Challenges and changes in gendered poverty: The feminization, de-feminization, and re-feminization of poverty in Latin America*. *Feminist Economics*, 2019. **25**(1): p. 119-144.
42. Fuentes, L., *(Re) reading the boundaries and bodies of femicide: exploring articulations within the discursive economy of gendered violence in 'post war' Guatemala*. 2016, Birkbeck, University of London.