URBAN INSTITUTE

DO NO HARM GUIDE

GLOBAL PERSPECTIVES ON EQUITY IN RESEARCH AND DATA

EDITED BY JONATHAN SCHWABISH, WESLEY JENKINS, AND SHUBHANGI KUMARI

OCTOBER 2024

TABLE OF CONTENTS

Introduction

Introduction

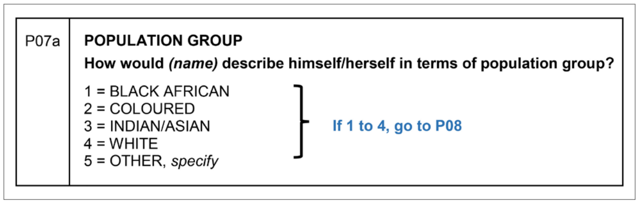
Context plays a crucial role in shaping research agendas, priorities, and methods. Contextual factors, such as the resources, history, and politics of a place or time, influence decisionmaking and can hinder research collaboration. At the Urban Institute, our work predominantly focuses on experiences within the United States, but we recognize that best practices for equitable data work in the US may not produce the same outcomes in international contexts. Therefore, for this seventh volume of the *Do No Harm Guide* series, *Global Perspectives on Equity in Research and Data*, we sought to learn more about best data practices from other parts of the world.

The motivation for creating this volume stems from a need to address global inequities, not just in economic growth or life expectancy but in the research and data ecosystems that affect how people live, work, and collaborate across contexts. While countries with fewer resources have often created cost-saving innovations in research and development, the inequitable distribution of research funding and knowledge can still perpetuate discriminatory outcomes by devaluing these achievements. In 2023, for example, the Indian Space Research Organization achieved the world's first landing on the south pole of the moon with a budget of $1.6 billion—or about 0.04 percent of the country's gross domestic product. By contrast, NASA's budget in 2023 was $25.4 billion, accounting for about 0.28 percent of the US gross domestic product and nearly 16 times its Indian counterpart.[1]

All around the world, research structures, tools, language, and analysis are shaped by the country's history and politics. South Africa's census survey has five categories for race, including "Coloured" (**figure 1**). Even though the term was constructed during apartheid to segregate people based on race, people in some South African communities or with certain backgrounds still use the term in a nondiscriminatory way.[2] The US census survey used the phrase "free colored persons" as a racial category in 1820, 1830, and 1840 (**figure 2**), but since then has used some combination of "Black," "Negro," and/or "African American."[3]

FIGURE 1

Race Categories from the 2022 South Africa Census Survey



P07a

POPULATION GROUP

How would *(name)* describe himself/herself in terms of population group?

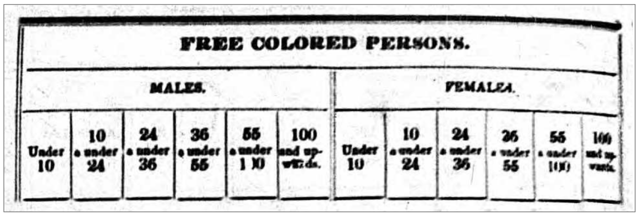
1 = BLACK AFRICAN
2 = COLOURED
3 = INDIAN/ASIAN
4 = WHITE
5 = OTHER, *specify*

If 1 to 4, go to P08

Source: Statistics South Africa, "Census 2022 CAPI/CATI Household Questionnaire," November 24, 2021, https://census.statssa.gov.za/assets/documents/2022/Census_2022_CAPI_CATI_Household_questionnaire.pdf.

FIGURE 2

"Free Colored Persons" Category from the 1840 US Census Survey



FREE COLORED PERSONS.											
MALES.						FEMALES.					
Under 10	10 & under 24	24 & under 36	36 & under 55	55 & under 100	100 and upwards.	Under 10	10 & under 24	24 & under 36	36 & under 55	55 & under 100	100 and upwards.

Source: National Archives and Records Administration, "SCHEDULE of the Whole Number of Persons within the Division Allotted to …," accessed February 23, 2024, https://www.census.gov/history/pdf/1840-1-042018.pdf.

Differences in how data are collected shape research, policies, and how people view themselves, their country, and the world. Our aim in this volume is to expand how researchers think about knowledge creation by highlighting adaptations and innovations in research processes and outcomes from around the globe. Such a worldview informs not just how US contexts shape research but how other countries adapt best practices into their unique contexts.

This volume contains four essays that broaden the conversations about context, from discussions of researcher bias, positionality, objectivity, and participatory methods—all of which have been discussed in previous volumes—to those that reflect on and critically analyze the global power dynamics prevalent in research. In these essays, the authors ask researchers to rethink their priorities, languages, methods, and outcomes from a more global perspective.

The volume begins with an essay by Malvika Sharan, a senior researcher at the Alan Turing Institute in London, UK, who dives into open science's economic and scientific divide and how to approach the pitfalls of open science using the "do no harm" principle. Anthea Piong, Andrea Berardo, and Murali Padmanabhan from the World Food Programme share how researchers can create cross-national, standardized datasets that would allow them to better understand the relationship between disability and food insecurity in the context of humanitarian operations at the World Food Programme. Julia Biedry Gonzalez, a technologist and data practitioner, illustrates the historical and modern data inequities of Puerto Rico in US federal products and how Hurricane Maria laid bare those disparities. Finally, Utami Diah Kusumawati and Yearry Panji Setianto from the journalism department at the Universitas Multimedia Nusantara in Indonesia discuss the growing and challenging state of data journalism education, data literacy, and data innovation in Indonesia.

We recognize that four essays by no means capture the breadth and variation of the historical and current challenges that data practitioners, researchers, and analysts around the world face when it comes to collecting, analyzing, and communicating data. Furthermore, we acknowledge that this volume does not incorporate the views of people with lived experience in how data has affected them, their families, or their communities. Importantly, we also know that research at the Urban Institute is conducted through a US-centric lens with a particular viewpoint—either implicit or explicit—on how data are and should be collected. Yet, even with these deficiencies and gaps, in an effort to move the field of equitable and inclusive data work forward, we think this guide can build on a conversation to bring together voices from across the globe.

In collecting and editing the essays for this volume, we hope to highlight the promise and pitfalls of research agendas and methods in a global context and how the data field can adapt to address inequities where needed. As researchers collaborate across borders to pursue equity in research processes and outcomes, it is imperative that we understand and challenge the forces that create scientific silos. The workarounds and innovations made in response to such forces presented in this volume offer many lessons on how researchers can improve their work, both at home and abroad.

NOTES

1. Michael Sheetz, "India's Moon Landing Made History at a Low Cost," CNBC, August 23, 2023, https://www.cnbc.com/2023/08/23/india-chandrayaan-3-moon-landing-came-at-small-cost.html.
2. "Diversity Terms in South Africa," *Profiles in Diversity Journal*, June 11, 2023, https://diversityjournal.com/10806-diversity-terms-in-south-africa/; and "Is the Word 'Coloured' Offensive When Writing about Apartheid?" Stack Exchange, accessed February 23, 2024, https://writing.stackexchange.com/questions/55141/is-the-word-coloured-offensive-when-writing-about-apartheid.
3. Pew Research Center, "What Census Calls Us: A Historical Timeline," accessed February 23, 2024, https://www.pewresearch.org/wp-content/uploads/2020/02/PH_15.06.11_MultiRacial-Timeline.pdf.

Chapter One

Aligning Open Science with "Do No Harm"

CHAPTER ONE

Aligning Open Science with "Do No Harm"

MALVIKA SHARAN

Open science—an umbrella term that refers to practices aiming to remove barriers to sharing, accessing, and reusing scientific research—seeks to promote commitments to ethical, equitable, and inclusive scientific approaches at all stages of research. Open-source software, open data standards, open education, citizen science, and open access are among a broad range of practices in open science that enhance the diversity of knowledge and knowledge producers (UNESCO 2018, 2021). Across different sectors and disciplines, open science is the common thread that defines scholarship and advances the goal of contributing to the production of knowledge as a public good (Arza and Fressoli 2017). Open science encourages greater collaboration across disciplines, ensuring that data are transparently shared and independently scrutinized by researchers, policymakers, and all others involved with or impacted by research and research outcomes.

The transparency inherent to open science practices can allow other researchers and the public to situate the work within a broader societal context, enabling everyone to evaluate research's potential benefits and harms. Unfortunately, the global research landscape today reflects our existing stratification, which limits the possibility for such scrutiny to happen. In 2022, total global research and development expenditures increased to $2.47 trillion, but just 10 high-income countries (HICs) accounted for about 85 percent of those expenditures (Sargent 2022).

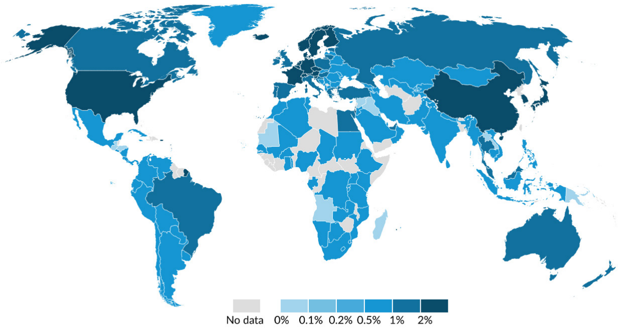
Positionality Statement

Malvika Sharan (she/her) is a cisgender Indian woman who grew up in Jharkhand—a multicultural, multireligious, and indigenous state in India that has a long-standing history of tribal struggle for representation and statehood, which was granted only two decades ago. Sharan's journey from North to South India, and later to Germany and the United Kingdom through economic immigration, marked by uncertainties and varying degrees of access, has strengthened her commitment to embedding diversity, equity, inclusion, and community approaches into research. Her background and identity situated within her country's colonial past, her home state's rich history, and her current privileges gained from working in Europe have profoundly influenced her research and capacity-building work in open science.

Low- and middle-income countries (LMICs), which make up approximately 70 percent of countries worldwide, accounted for less than 1 percent of research and development expenditures (Okoye et al. 2022; **figure 1**).

FIGURE 1

Research and Development Spending as a Share of GDP, 2021



Source: “Research & Development Spending as a Share of GDP, 2021,” Our World in Data, accessed January 11, 2024, https://ourworldindata.org/grapher/research-spending-gdp. Data are from UNESCO via the World Bank.

Note: Spending includes current and capital expenditures (public and private) on research.

The economic and scientific divide between HICs and LMICs—terms often used for the global North and global South (henceforth “global majority”)[1]—is both a cause and consequence of the global disparity in research and development. The lack of investment in research infrastructure and significantly fewer researchers in secure, professional roles in LMICs continue to widen this divide.[2] While open science could make the scientific process more transparent, inclusive, and democratic, it has yet to offset the widespread inequity in how knowledge and economic benefits are distributed in research and how unequal collaborative structures continue to exclude already marginalized communities. As such, we must ask, how can open science further equity in the international research ecosystem? Can open science facilitate a research culture where diverse people participate and are recognized, not just in accessing knowledge but in its production, circulation, and societal benefit? How can open science “do no harm”?

In this essay, I explore how “do no harm,” which is meant to help organizations become “more effective, accountable, and efficient” by considering the context within which an intervention is applied, could advance the goals of open science and tackle widespread inequity in the research ecosystem (CDA 2018). Specifically, I highlight how open science grapples with the entrenched power and resource imbalances between HICs and LMICs that disproportionately benefit a handful of HIC actors over others. In exploring these intended roles and unintended consequences of open science in research, I connect them with “do no harm” principles that can help meaningfully achieve open science’s goal of “uniting humanity in a common intellectual conversation and quest for knowledge,” a definition created by the Budapest Open Access Initiative.[3]

Confronting Inequity in the Research Ecosystem

When thousands of individuals and organizations worldwide signed the Budapest Open Access Initiative declaration, they committed to pursuing open-access publications in research and science. The hope was that removing access barriers to academic literature would accelerate research and “share the learning of the rich with the poor and the poor with the rich.”[4] Although this statement paints a hopeful picture, it is sadly far from reality. While open science efforts have undoubtedly enabled the collaboration and sharing of research outcomes, and in many cases fair recognition through authorship and credits with everyone involved, wide global disparities remain.

Open science is necessary to promote global collaboration beyond the confines of the global North to enrich and diversify our understanding and innovate together as an international scientific community. Global majority-led cooperation and collaborations, for instance, have been instrumental in advancing knowledge, particularly in areas such as health and climate research.[5] But the research community has yet to dismantle the subtle power imbalances and social subordination in the global research ecosystem. North–global majority collaborations can often have hegemonic influences, even when a research project is global majority–led.[6] Researchers working in LMICs often do not have the same economic opportunity or power as those working in HICs to direct research in the interest of their local communities.

The UK and US hegemony in science also means that most publications and resources are expected to be published in English. Researchers whose primary language is not English are required to invest significant amount of time and effort into translating technical information from their native language to English to serve international communities. In contrast, adoption of international practices and learning materials requires translating and contextualizing them to reflect local cultures, experiences, and resource availability. Open science means that more people around the world can access different resources, yet researchers from regions that are already underfunded and often unrecognized face an excess burden.

Most care work in research—for example, embedding considerations for digital accessibility, disability, and neurodiversity, as well as general community building—remain a hidden aspect of open science. This care infrastructure has largely depended on volunteer labor, frequently carried out by people with minority identities and backgrounds (D'Ignazio and Klein 2020). They tend to take on these "hidden labor" costs, often without being fairly credited or supported for their work (Hostler 2023).

As open science becomes standard practice across research, it is important to identify, improve awareness of, and reduce its potential negative effects. By introducing a "do no harm" framework, open science can better account for its potential harms and reduce global disparities in research production.

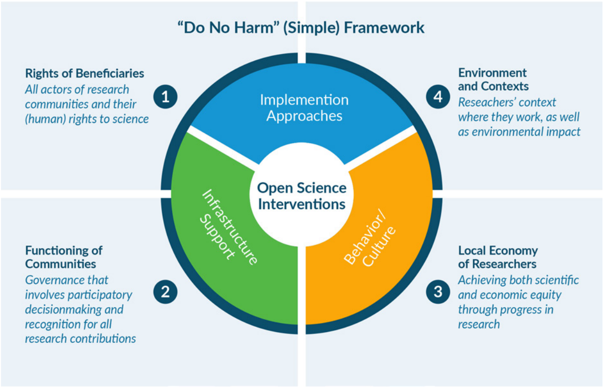
A Framework for Aligning Open Science with "Do No Harm"

The "do no harm" principle has been widely applied in medical research and health care,[7] and it was later extended into humanitarian aid work to consider both intended and unintended consequences of an intervention (CDA 2018). Many institutional policies, ethical frameworks, and research guidelines also mention or implicitly incorporate the principle. In a technical context, the Organization for Ethical Source is in the early stages of developing ethical source licenses (e.g., Hippocratic and Do No Harm licenses) to promote responsible use of software.[8] However, "do no harm" has not been clearly discussed in the larger context of open science.

Figure 2 illustrates a simplified framework through which open science can be aligned with the "do no harm" principle by considering four separate groups: rights of beneficiaries, community functioning and relationship between actors, the local economy of researchers, and the broader environment and contexts.

FIGURE 2

The Simplified "Do No Harm" Framework



Source: Malvika Sharan, "Exploring 'Do No Harm' Principles in Open Research Communities," presentation from keynote talk at the 2nd Funders Workshop, Montreal, Canada, September 20, 2023, https://doi.org/10.5281/zenodo.8361334. Adapted from Jean Martial Bonis Charancle and Elena Lucchi, *Incorporating the Principle of "Do No Harm": How to Take Action without Causing Harm* (Silver Spring, MD: Humanity & Inclusion; Paris: F3E, 2018).

Encompassing four distinct aspects of the current research ecosystem, this framework can be a useful guide for researchers to identify and minimize the potential negative effects of open science in their research.

① Rights of Beneficiaries

All actors of research communities and their (human) rights to science

Transparency and openness in research improve access to information and enable better replication and reproduction efforts, accelerating scientific advancement across disciplines. The global response to the COVID-19 pandemic showed just how effective open practice can be. The pandemic compelled governments, funders, researchers, and members of the scientific community to come together and gather resources, create systems of knowledge exchange, and prioritize human lives and public safety. Despite this global collaboration, the result still reflected existing inequities—as of November 2023, more than 79 percent of people in HICs are vaccinated for COVID-19, compared with only 32 percent of people in low-income countries.[9] Vaccine nationalism, intellectual property rights, constraints in manufacturing capacity, poor resilience of health care systems, and vaccine hesitancy are among many factors that have affected the potential of LMICs to fully vaccinate their citizens (Ferranna 2023). In this global effort, UNESCO (the United Nations Educational, Scientific and Cultural Organization) mobilized more than 122 HICs and LMICs to promote open science and reinforced international cooperation to lift patents for vaccine equity.[10]

This story, among many such examples of worldwide collaboration, shows how open science can advance equity by bringing together international communities of experts to address common challenges. The translation of collaboration into equitable benefits, however, is not guaranteed. Open practices during the COVID-19 pandemic allowed research findings to be shared rapidly, such as through preprint publications, crowdsourcing, and citizen involvement. But the urgency to gain scientific information and publish data also introduced significant challenges, including the dissemination of low-quality research methods, unverified scientific claims, and fabricated data. As a result, harmful outcomes took hold, such as the administration of unproven medical products, poorly designed guidelines, disinformation related to COVID-19 cures, and biased tools that especially affected vulnerable communities (Harper et al. 2020).

Open ways of working should not be considered a protocol to use only in times of crisis. Instead, they should be implemented in research processes and ecosystems to ensure that the rights of the groups most marginalized from benefits are at the forefront of any decisionmaking. For this vision to become a reality, doing no harm must be considered in the implementation of open practices,[11] with the understanding that those practices are for the benefit of all research stakeholders (UNESCO 2018). For any project, practitioners should openly ask, who makes decisions? Who gets to participate in knowledge production (Chan and Mounier 2019)? Who controls and protects the system? Who enjoys and benefits from the outcome? And who is responsible for addressing challenges and harm?

"Do no harm" requires researchers and organizations to acknowledge power dynamics that participants and beneficiaries may face, as well as potential biases and assumptions that may affect the outcomes of their research. In combination with open science, "do no harm" can necessitate that researchers prioritize research agendas with the involvement and engagement of individuals and communities who may be affected by their research. It can also guide the creation of inclusive, respectful, and collaborative environments where these dynamics and assumptions can be openly discussed and collectively dismantled.

② Functioning of Communities

Governance that involves participatory decisionmaking and recognition for all research contributions

For any kind of community to thrive, it needs a system of governance—that is, a set of formal and informal practices through which a group can set goals, assign responsibilities, establish systems, and assess outcomes of organizational action (Grandori 2013). In open science projects, transparent governance provides structure and a process through which all participants and beneficiaries can influence decisionmaking. In the context of research and data science, governance provides direction, support, and accountability, so that outcomes are co-created to benefit all members. A governance structure that embodies open principles sets the tone for an inclusive research culture that fosters collaboration and the active involvement of diverse actors in the decisionmaking process.

In the context of doing no harm, open practices can be implemented to democratize access to public goods. Consider the "tragedy of the commons" argument popularized by biologist Garrett Hardin (1968): "Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the common." About thirty years later, in her book *Governing the Commons*, Elinor Ostrom exposed the limitations of Hardin's theory by arguing that Hardin

- focused on "access" without "governance,"
- assumed little or no communication between the people involved,
- postulated that people act only in their immediate self-interest, and
- offered only two solutions to correct the tragedy, privatization or government intervention (Ostrom 1990).

Hardin also emphasized a "pasture" open to "all men" operating in a world without intersectional perspectives, solidarity, or collective actions. Ostrom's theory of governance of the commons, by contrast, explains that doing no harm goes beyond simple access; it requires open practices in governance for the democratization of shared resources. In open science, resources include funding, infrastructure, and support at individual, institutional, geographical, and political levels. Good governance can ensure better community functioning and relationships by democratizing co-creation, regulation, and sharing of resources through open collaboration.

The positive functioning of communities inherently requires open communication and transparency around how decisions are made, how different members or representative groups are involved, and how those decisions affect the community. These considerations in governance can empower participants and help redress power imbalances, such as those in global majority–global North research collaborations.

The "Do no harm" principle in community-based governance of open science practices can allow fair regulation of research infrastructures and collaborative solutions that contribute to equitable participation. When working with communities, "do no harm" informs the use of culturally appropriate research methodologies, ensures informed consent, and establishes pathways to integrate the involved communities' perspectives into the research outputs. Through involvement of different stakeholders in decisionmaking, it can also help prioritize the use of infrastructure, such as for ensuring accessibility, providing financial support, offering upskilling opportunities, and implementing policies appropriately to promote openness and equity in research.

③ Local Economy of Researchers

Achieving both scientific and economic equity through progress in research

Local economies in research are defined as resources, infrastructure, and platforms available for researchers and communities to address local needs. In the global research landscape, only a small proportion of stakeholders—such as large, powerful institutions and commercial entities in academic publishing—have profited financially and reputationally from closed systems of knowledge production and circulation. These groups, largely in HICs, have also resisted and often obstructed the open science movement to protect their dominant status as publishers, infrastructure providers, and regulatory bodies (Chan and Mounier 2019).

Even when these select groups show support for open-access and open-research practices, their profit-making business models continue to undermine the complexity and nuance in research conducted in LMICs. For example, international funders and philanthropic organizations working to advance open science recognize the challenges of limited investment and funding options in LMICs and tend to support collaborations between HICs and LMICs. However, they often end up funding only the "competitive" résumés and proposals, intentionally or unintentionally, even though what makes a proposal competitive is defined by a narrow Western-centric perspective (Reidpath and Allotey 2019). Historically, this bias has largely favored researchers and organizations both in HICs and LMICs who can

- navigate the funding landscape and pursue research priorities relevant to HICs,
- collaborate with HIC institutions that set the research agenda, or
- provide "competitive" scientific service beyond their core research goals.

Instead of empowering research conducted in LMICs to benefit from open access, the current academic incentive model only serves those who can afford to participate in their system,[12] while labeling those seeking to benefit as "free riders." Such simplistic, reductionist, and partial view of open science has advanced an anticommons or antitrust agenda, which has fueled excessive intellectual property rights and over-patenting, leading to privatization, vendor lock-in, commercial exploitation, and underuse of scientific resources (Lieberwitz 2005).

At the heart of this profit-making business model is an antiquated reward system that locks in the economic and academic capital of those who already benefit and disadvantages the possibility of equity for LMIC researchers (Sharan and MacCallum 2023). High journal subscription fees, steep article processing charges, costly software and hardware required to carry out specialized research, and expensive membership fees to access conferences where "important" research is presented are some of the high barriers to participation for researchers from LMICs. Privileged stakeholders have been self-selected to construct and impose unreasonable measures of success, ensuring they continue to benefit from the system by maintaining such high access and contribution costs. These challenges are exacerbated by a lack of local infrastructure, which is needed to build and adequately support the global scientific workforce (Aschauer 1989; Straub 2008; Whittle 2009).

In the context of the local economy of researchers, the "do no harm" principle should seek to uphold people's rights to access and participate in science to meet their communities' needs.[13] Alignment of open science with "do no harm," therefore, should go beyond maintaining open collaboration with diverse members. It should also require the identification and mitigation of potential economic risks associated with a research initiative, provide ways for local communities to participate, and create new economic opportunities for them.

"Do no harm" enables deeper respect and appreciation for the cultural and social dynamics of the local community, avoiding any unintended economic disruptions or social tensions that can result from an intervention. By following participatory research approaches such as asset-based community development (Vaughn and Jacquez 2020), researchers can embed local knowledge, build on the strengths of local communities, and bring back scientific and economic benefits to them. This change in research practices will increase public trust, develop norms of reciprocity, and evolve academic systems through the involvement of a more diverse set of stakeholders.

④ Environment and Contexts

Reseachers' context where they work, as well as environmental impact

In an era with myriad global challenges—including climate change, natural disasters, health crises, and violent conflicts—the affected communities need to be valued as much as the research and data itself. Open science can be much more effective when aligned with the "do no harm" principle, leading to more socially responsible research processes, outputs, and impacts than when applied on its own. In this final part of the framework, I explore how the "do no harm" principle can promote open science in ways that do not harm but enhance the research environment. The term "environment" here is used to describe the local contexts of communities, the natural environment affected by research, and the work environment of researchers. Practices that respect the rights and knowledge systems of local and Indigenous communities, safeguard biodiversity within the area of intervention, and protect the working conditions of people offer some of the most well-established ways to protect environments (Lent 2019). Here, I highlight some of those practices that open science researchers can adopt.

Local contexts of communities: Open science has led to radical knowledge sharing and interdisciplinary collaboration in research to maximize the use of big data and data-informed technologies that address the challenges faced by society (Leonelli 2020). Recent advances in artificial intelligence (AI) have led to ethical discourse around the benefits and harms of data, as well as the democratization of open source AI technology that responds to local needs (Buolamwini 2023; Gwagwa et al. 2024; O'Neil 2017). When working with data, it is vital to embed the FAIR (Findable, Accessible, Interoperable, Reusable) principles for data management, while making information "as open as possible, as closed as necessary" (European Commission 2016, 4). These principles, along with its open source tools and reproducibility frameworks, can help ensure that health, location, and other identifiable data from local and Indigenous communities are guarded against exploitation while advancing the transparency of research methods, processes, and purposes. The FAIR principles can also help make digital assets other than data findable, accessible, interoperable, and reusable while maintaining privacy and protection where necessary.[14] Critically, these principles are encouraged to be applied along with the CARE Principles for Indigenous Data Governance developed by the Research Data Alliance's International Indigenous Data Sovereignty Interest Group to address power imbalances in who gets to decide what is "necessary."[15] The CARE principles—collective benefit, authority to control, responsibility, and ethics—demand the decolonization of research practices and technology and the safeguarding of vulnerable communities by protecting their knowledge, such as land data and Indigenous languages, against exploitation.[16]

Natural environment affected by research: Open collaboration has led to many breakthroughs in research by applying big-team science, multilingual, and cross-regional models that contextualize practices and consider the local effects of research, including environmental impact. More recently, open science has informed practices for building technologies that minimize environmental harm and embed climate justice at different stages of research. AI technology and open-source tools have been developed to actively "measure, optimise and reduce" carbon emissions.[17] In addition, several research-based recommendations have been put forward to encourage collaboration between state and local actors to support the adoption of open-source solutions while improving incentives and financial investment structures in research for the benefit of the environment (Augspurger, Malliaraki, and Hopkins 2023; Govaart, Hofmann, and Medawar 2022).

Work environment of researchers: Commitment to socially responsible and DEI (Diversity, Equity and Inclusion) practices in the workplace has been extensively researched, and numerous studies have connected fairness, safety, and employment conditions to the well-being of employees.[18] Although not unique to research, open practices can contribute to improving transparency, accountability, and participatory approaches, which leads to ethical working conditions, employment rights, and academic freedom of researchers (Vrielink, Lemmens, and Parmentier 2011). Furthermore, equitable opportunities, recognition, and fair compensation for all contributors to research, along with open governance approaches, can significantly improve the work environment in research (Buchanan and Warwick 2021; Zhenjing et al. 2022).

Arguably, this aspect of “do no harm” in open science encompasses and depends on the practices described for the other three aspects of the framework. It also aligns most closely with the “do no harm” principle as understood in humanitarian aid work, which aims to minimize negative effects on the environment by considering the potential consequences of human actions on surrounding ecosystems. Considering open science approaches as interventions, lessons from humanitarian frameworks can be directly translated into the research ecosystem. The “do no harm” principle in this context can help embed holistic approaches to assess and monitor the implementation of open science with a goal to protect local communities, their natural environment, and the working conditions of all contributors of research against exploitation and harm.

Conclusion

Openness has proven itself time and again as a facilitator for responsible research outcomes that are independently scrutinized, reused, and improved. Although open science has significantly refined research practices to increase participation and access to knowledge systems, poorly designed open science approaches can have negative effects. As described in this essay, a “do no harm” framework can assess the benefits and harms of open science. Taking its aspects as starting points, researchers, institutions, and local communities can identify and mitigate negative effects by improving open science implementation, providing infrastructure support, and promoting behaviors for positive research culture.

“Do no harm” in open science calls for reexamining privilege, access, and benchmarks for academic success and research impact. It should be taken as an additional layer of consideration, not as a restriction or distraction. It is a guide for researchers on how to work in solidarity with others and the environment where they conduct their research. With this safeguarding, open science can help create a more equitable space, where diverse actors with wide range of backgrounds, identities, abilities, expertise, and knowledge are valued and given the opportunities to participate fully in the global research economy—achieving knowledge equity, sharing economic benefits, and reducing harm.

NOTES

1. "Global majority" refers "to people who are racialized as Black, African, Asian, Brown, dual-heritage, indigenous to the Global South and/or … 'ethnic minorities.' These groups currently represent approximately 80% of the world's population" ("People of the Global Majority," Immigration Law Practitioners' Association, accessed July 23, 2024, https://ilpa.org.uk/people-of-the-global-majority/). See also Campbell-Stevens (2021). "South" is used occasionally in this essay in phrases like "South-South cooperation."
2. "Research and Development Statistics (RDS)," Organisation for Economic Co-operation and Development, accessed January 11, 2024, http://oe.cd/rds.
3. "Read the Declaration," Budapest Open Access Initiative, accessed July 17, 2024, https://www.budapestopenaccessinitiative.org/read/. The Budapest Open Access Initiative was launched in 2002.
4. "Read the Declaration," Budapest Open Access Initiative.
5. Dima Al-Khatib, "South-South Cooperation Is Essential to Achieving the Sustainable Development Goals," *UN Chronicle*, United Nations, September 11, 2023, https://www.un.org/tr/node/209190.
6. Sophie Inge, "Researchers 'Must Recognise North-South Power Imbalance,'" Research Professional News, June 15, 2023, https://www.researchprofessionalnews.com/rr-news-uk-universities-2023-6-researchers-must-recognise-north-south-power-imbalance/; and "On South-South Cooperation: Assessing Its Political Relevance and Envisioning a Future Beyond Technical Cooperation," The Reality of Aid Network, October 26, 2023, https://realityofaid.org/reality-check-ssc/.
7. "Research Must Do No Harm: New Guidance Addresses All Studies Relating to People," Editorial, *Nature*, June 14, 2022, https://doi.org/10.1038/d41586-022-01607-0.
8. "Ethical Source Licenses," Organization for Ethical Source, accessed February 6, 2024, https://ethicalsource.dev/licenses/; and "The Hippocratic License," Organization for Ethical Source, accessed February 6, 2024, https://firstdonoharm.dev/.
9. "Global Dashboard for Vaccine Equity," United Nations Development Programme, accessed July 30, 2024, https://data.undp.org/insights/vaccine-equity.
10. Memo from Audrey Azoulay (Director-General, UNESCO) to ministers responsible for relations with UNESCO, "Subject: Preliminary Report on the first draft of the Recommendation on Open Science," June 2023, https://webarchive.unesco.org/20230617030159/https://en.unesco.org/sites/default/files/comments_osr_slovakia_document.pdf; "The Netherlands' response to the first draft of the UNESCO Recommendation on Open Science," June 2023, https://webarchive.unesco.org/20230617025504/https://en.unesco.org/sites/default/files/comments_osr_netherlands_document.pdf; and UNESCO, "UNESCO Calls for COVID-19 Vaccines to Be Considered a Global Public Good," press release, last updated April 20, 2023, https://www.unesco.org/en/articles/unesco-calls-covid-19-vaccines-be-considered-global-public-good.
11. Laura Carter, "The Human Rights Case for Open Science," Open Heroines, August 10, 2022, https://openheroines.org/blog/the-human-rights-case-for-open-science.
12. Kaitlin Thaney, "Open Access 'at Any Cost' Cannot Support Scholarly Publishing Communities," *LSE Impact Blog*, London School of Economics and Political Science, July 20, 2023, https://blogs.lse.ac.uk/impactofsocialsciences/2023/07/20/open-access-at-any-cost-cannot-support-scholarly-publishing-communities.
13. See, for example, United Nations Office of the High Commissioner for Human Rights, "The Right to Access to and Participate in Science," accessed July 30, 2024, https://www.ohchr.org/en/special-procedures/sr-cultural-rights/right-access-and-participate-science.
14. "FAIR Principles," GO FAIR, accessed January 15, 2024, https://www.go-fair.org/fair-principles.
15. Stephanie Russo Carroll, Maui Hudson, Jarita Holbrook, Simeon Materechera, and Jane Anderson, "Working with the CARE Principles: Operationalising Indigenous Data Governance," Ada Lovelace Institute blog, November 9, 2020, https://www.adalovelaceinstitute.org/blog/care-principles-operationalising-indigenous-data-governance.
16. Kimberley Brown, "Indigenous Lands, Knowledge Are Essential for Saving Primates from Extinction, Says New Study," Mongabay, August 11, 2022, https://news.mongabay.com/2022/08/indigenous-lands-and-knowledge-are-essential-for-saving-primates-from-extinction-says-new-study/; and Keoni Mahelona, Gianna Leoni, Suzanne Duncan, and Miles Thompson, "OpenAI's Whisper Is Another Case Study in Colonisation," Papa Reo blog, January 24, 2023, https://blog.papareo.nz/whisper-is-another-case-study-in-colonisation/.
17. Tobias Augspurger, "4 Ways You Can Preserve the Earth's Livability with Open Source," Opensource.com, May 17, 2022, https://opensource.com/article/22/5/preserve-earths-livability-open-source.
18. Brian Kropp, Jessica Knight, and Jonah Shepp, "How Fair Is Your Workplace?" *Harvard Business Review*, July 14, 2022, https://hbr.org/2022/07/how-fair-is-your-workplace; and "Fairness and Wellbeing in the Workplace," University of Manchester, Work and Equalities Institute, accessed August 27, 2024, https://www.wei.manchester.ac.uk/research/themes/fair-treatment-at-work.

REFERENCES

Arza, Valeria, and Mariano Fressoli. 2017. "Systematizing Benefits of Open Science Practices." *Information Services & Use* 37 (4): 463–74. https://doi.org/10.3233/ISU-170861.

Aschauer, David A. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23:177–200, cited in Straub (2008).

Augspurger, Tobias, Eirini Malliaraki, and Josh Hopkins. 2023. *Open Source in Environmental Sustainability*. Geneva, CHE: Zenodo. https://doi.org/10.5281/zenodo.7771633.

Bonis Charancle, Jean Martial, and Elena Lucchi. 2018. *Incorporating the Principle of "Do No Harm": How to Take Action without Causing Harm*. Silver Spring, MD: Humanity & Inclusion; Paris: F3E.

Buchanan, David, and Ian Warwick. 2021. "First Do No Harm: Using 'Ethical Triage' to Minimise Causing Harm When Undertaking Educational Research among Vulnerable Participants." *Journal of Further and Higher Education* 45 (8): 1090–103. https://doi.org/10.1080/0309877X.2021.1890702.

Buolamwini, Joy. 2023. *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. New York: Random House.

Campbell-Stephens, Rosemary M. 2021. "Introduction: Global Majority Decolonising Narratives." In *Educational Leadership and the Global Majority*, 1–21. Cham, CHE: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-88282-2_1

CDA (Collaborative for Development Action). 2018. "Do No Harm: A Brief Introduction from CDA." Cambridge, MA: CDA Collaborative Learning Projects.

Chan, Leslie, and Pierre Mounier, eds. 2019. *Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure*. Marseille, FRA: OpenEdition Press. https://doi.org/10.4000/books.oep.8999.

D'Ignazio, Catherine, and Lauren F. Klein. 2020. "7. Show Your Work." In *Data Feminism*, 173–203. Cambridge, MA: MIT Press.

European Commission. 2016. *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*. Brussels: European Commission, Directorate-General for Research and Innovation.

Ferranna, Maddalena. 2024. "Causes and Costs of Global COVID-19 Vaccine Inequity." *Seminars in Immunopathology* 45:469–80. https://doi.org/10.1007/s00281-023-00998-0.

Govaart, Gisela H., Simon M. Hofmann, and Evelyn Medawar. 2022. "The Sustainability Argument for Open Science." *Collabra: Psychology* 8 (1): 35903. https://doi.org/10.1525/collabra.35903.

Grandori, Anna, ed. 2013. *Handbook of Economic Organization*. Cheltenham, UK: Edward Elgar Publishing.

Gwagwa, Arthur, Brandon Jackson, Elliot Jones, Stacey King, Mira Lane, Micaela Mantegna, Thomas Schneider, Kathleen Siminyu, and Alek Tarkowski. 2024. *Artificial Intelligence and the Challenge for Global Governance*, edited by Alex Krasodomski. London: Chatham House. https://doi.org/10.55317/9781784136086.

Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162 (3859): 1243–48. https://doi.org/10.1126/science.162.3859.1243.

Harper, L., N. Kalfa, G. M. A. Beckers, M. Kaefer, A. J. Nieuwhof-Leppink, Magdalena Fossum, K. W. Herbst, and D. Bagli. 2020. "The Impact of COVID-19 on Research." *Journal of Pediatric Urology* 16 (5): 715–16. https://doi.org/10.1016/j.jpurol.2020.07.002.

Hostler, Thomas J. 2023. "The Invisible Workload of Open Research." *Journal of Trial & Error* 4 (1). https://doi.org/10.36850/mr5.

Lent, Marina. 2019. "Sustainable Development: First, Do No Harm." In *Spotlight on Sustainable Development 2019*, edited by Barbara Adams, Cecilia Alemany Billorou, Roberto Bissio, Chee Yoke Ling, Kate Donald, Jens Martens, and Stefano Prato, 74–85. New York: Civil Society Reflection Group on the 2030 Agenda for Sustainable Development.

Leonelli, Sabina. 2020. "Scientific Research and Big Data." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford, CA: Metaphysics Research Lab, Stanford University.

Lieberwitz, Risa L. 2005. "Confronting the Privatization and Commercialization of Academic Research: An Analysis of Social Implications at the Local, National, and Global Levels." *Indiana Journal of Global Legal Studies* 12 (1): 109–52. https://doi.org/10.1353/gls.2005.0005.

Okoye, Kingsley, Julius T. Nganji, Jose Escamilla, Jin Michael Fung, and Samira Hosseini. 2022. "Impact of Global Government Investment on Education and Research Development: A Comparative Analysis and Demystifying the Science, Technology, Innovation, and Education Conundrum." *Global Transitions* 4:11–27. https://doi.org/10.1016/j.glt.2022.10.001.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Reidpath, Daniel D., and Pascale Allotey. 2019. "The Problem of 'Trickle-Down Science' from the Global North to the Global South." *BMJ Global Health* 4:e001719. https://doi.org/10.1136/bmjgh-2019-001719.

Sargent, John F., Jr. 2022. "Global Research and Development Expenditures: Fact Sheet." CRS R44283. Washington, DC: Congressional Research Service.

Sharan, Malvika. 2023. "Exploring 'Do No Harm' Principles in Open Research Communities." Presentation from keynote talk at the 2nd Funders Workshop, Montreal, CAN, September 20. https://doi.org/10.5281/zenodo.8361334.

Sharan, Malvika, and Catriona MacCallum. 2023. "Resolved: Open Practices Make Science Better—Debate Materials." Opening speech at the Researcher to Reader Conference, London, ENG, February 21–22. https://doi.org/10.5281/zenodo.7687846.

Straub, Stéphane. 2008. "Infrastructure and Growth in Developing Countries: Recent Advances and Research Challenges." Policy Research Working Paper 4460. Washington, DC: World Bank. Available at SSRN: http://ssrn.com/abstract=1080475.

UNESCO (United Nations Educational, Scientific and Cultural Organization). 2018. *Recommendation on Science and Scientific Researchers*. Paris: UNESCO.

———. 2021. *UNESCO Recommendation on Open Science*. Paris: UNESCO.

Vaughn, Lisa M., and Farrah Jacquez. 2020. "Participatory Research Methods–Choice Points in the Research Process." *Journal of Participatory Research Methods* 1 (1). https://doi.org/10.35844/001c.13244.

Vrielink, Jogchum, Paul Lemmens, and Stephan Parmentier. 2011. "Academic Freedom as a Fundamental Right." *Procedia – Social and Behavioral Sciences* 13 (2011): 117–41. https://doi.org/10.1016/j.sbspro.2011.03.009.

Whittle, James. 2009. "TI-UP Report – A Short Synthesis of the Link between Infrastructure Provision / Adequacy and Economic Growth." Report prepared for the TI-UP Resource Centre. London: Department for International Development.

Zhenjing, Gu, Supat Chupradit, Kuo Yen Ku, Abdelmohsen A. Nassani, and Mohamed Haffar. 2022. "Impact of Employees' Workplace Environment on Employees' Performance: A Multi-Mediation Model." *Frontiers in Public Health* 10. https://doi.org/10.3389/fpubh.2022.890400.

Chapter Two

Disability Data in Humanitarian Contexts

CHAPTER TWO

Disability Data in Humanitarian Contexts

ANTHEA PIONG

ANDREA BERARDO

MURALI PADMANABHAN

In recent years, despite increased efforts to raise awareness of the needs of people with disabilities and to make their life more empowering and accessible, many spaces still remain difficult to navigate for people with disabilities. An important first step for creating a more diverse and inclusive society is to collect data about people with disabilities by asking questions like "Is this person disabled?" This question may seem simple, but it hides many layers of assumptions that lead to more questions: What does it mean to be disabled? What counts as a disability? Is disability a spectrum?

In 2001, the United Nations sponsored the International Seminar on the Measurement of Disability, where experts reviewed and assessed methods of collecting data about people with disabilities —and set a course on how to improve them.[1] The goal was to propose a new approach that would have "common definitions, concepts, standards, and methodologies" across the globe and would result in "internationally comparable, high-quality disability data collection" (WG 2020, 1). The outcome of those meetings was to form the Washington Group on Disability Statistics (WG) that would address the need for better and more consistent data. The Washington Group was tasked with developing standardized questions, methodologies, and guidelines to aid data collection (**box 1**).

Positionality Statements

Anthea Piong (she/her) is one of a handful of Singaporeans working for the World Food Programme. She started her career in Washington, DC, then entered the field of international development as a monitoring and evaluation officer in Sierra Leone through a Princeton in Africa fellowship, gaining valuable hands-on experience with collecting and using data to benefit the lives of rural farmers. She moved on to focus on humanitarian work with the World Food Programme in Bangkok. Piong currently works for WFP in the Afghanistan country office, working with data to inform evidence-based and inclusive program design for WFP beneficiaries.

Andrea Berardo (he/him) is an Italian man who grew up in a small town in the north of the country. Passionate about fighting poverty and inequality, he engaged first in academic research and later on in humanitarian and development work. He now leads a team of researchers at the World Food Programme Regional Bureau for Asia and the Pacific. Berardo treasures his experience working directly with people assisted by WFP, promoting dignity and inclusion as ways to identify the most suitable social interventions.

Murali Padmanabhan (he/him) is an inclusive development professional with more than 28 years of experience working among fellow people with disabilities to promote and protect their rights through advocacy, research, and guiding mainstream organizations toward inclusion for an equitable society. Belonging to the South Asian fraternity, Padmanabhan has had the opportunity to work with grassroots to international entities. He currently works at the WFP Bangkok in a multicultural environment.

The WG created sets of questions, or modules, for different populations and contexts (e.g., children, labor force participation, school participation, etc.),[2] and governments and organizations around the world have used the modules to collect data using a standardized, globally comparable methodology. One question set, the Washington Group Short Set on Functioning (WG-SS),[3] is most commonly used by humanitarian organizations like the United Nations World Food Programme (WFP) to collect data on the disability status of the people it reaches with life-saving assistance. However, the use of the WG-SS in these settings still poses some limitations in the collection of data on disability, as humanitarian work may not have been one of the contexts the WG originally considered in 2001. In this essay, we discuss some challenges the WFP faces in using WG-SS questions in the organization's humanitarian operations and propose an adaptation of the questions that could help minimize those challenges.[4]

BOX 1: THE WASHINGTON GROUP ON DISABILITY STATISTICS

The Washington Group on Disability Statistics (WG) was formed at the 2001 United Nations–sponsored International Seminar on Measurement of Disability. It is what the United Nations calls a "city group," which invites representatives from national statistical agencies across the world to address the "statistical challenges to collecting valid, reliable, and cross-nationally comparable data." This effort joins representatives from statistical agencies with various international agencies and other experts, including United Nations agencies such as UNICEF, the International Labour Organization, and the World Bank; bilateral aid agencies such as the United States Agency for International Development; nongovernment organizations; disabled people's organizations; and researchers. Since the convening, more than 135 countries have participated in the WG and more than 80 countries have used the proposed questions in their censuses or surveys.

For more information about the Washington Group on Disability Statistics, visit its website at https://www.washingtongroup-disability.com/.

About the WG-SS

The WG-SS consists of six questions for use in national censuses and surveys; it is a tool for disaggregating data for individuals with disabilities, not for diagnosing or identifying disability (WFP 2022). In other words, it is not intended to be used to collect data about specific types of disability or to prescribe solutions or programs to address an individual's disabilities. Instead, it is intended to identify whether a person is disabled and to provide information that can direct large-scale decisionmaking on policies or programs.

The questions are phrased to focus on six areas of difficulties a person may have performing basic activities: difficulties seeing, hearing, walking or climbing steps, remembering or concentrating, undertaking self-care, and communicating (expressive and receptive) (**box 2**). A respondent who answers at least one of the six questions with "A lot of difficulty" or "Cannot do at all" would be classified as a person with a disability at the time of the survey, while those who only answer with "Some difficulty" or "No difficulty" would not be classified as a person with a disability.

The questions are intentionally meant to be a respondent's subjective self-assessment, giving them dignity and authorship to decide for themselves how they experience a disability rather than imposing a definition on them that may be subject to the biases and interpretations of the interviewer. While the self-assessment approach has been shown to lead to undercounts of people with disability in research-focused contexts (Landes, Swenor, and Vaitsiakhovich 2024), the methodology is an improvement over other methods in its nuance and respects the respondent's self-identification.

Before the widespread use of the WG-SS, organizations like the WFP asked about disability status in a simplistic way—for example, "Is anyone in your house disabled? Yes/No,"—leaving it up to the respondent or interviewer to interpret what "disabled" means. In many countries, the term "disabled" can be stigmatizing, and the respondent can face discrimination from neighbors, exclusion from the community, or even lose income opportunities if perceived as "disabled" according to local customs and norms. These consequences are particularly relevant in situations where respondents may not have enough privacy while taking the survey. The WG-SS intends to counter this stigmatization by using "difficulty" rather than "disability," and any subjectivity is confined to the respondent's answer during their self-assessment, while the phrasing of the question itself remains consistent. For these reasons, the WG does not recommend using questions like the one WFP previously used.

BOX 2: THE WASHINGTON GROUP SHORT SET ON FUNCTIONING QUESTIONS

The following six questions ask about activities a person may have difficulty doing because of a health problem:

1. Do you have difficulty seeing, even if wearing glasses?
2. Do you have difficulty hearing, even if using a hearing aid?
3. Do you have difficulty walking or climbing steps?
4. Do you have difficulty remembering or concentrating?
5. Do you have difficulty with self-care, such as washing all over or dressing?
6. Using your usual (customary) language, do you have difficulty communicating, for example understanding or being understood?

Response options: No difficulty, Some difficulty, A lot of difficulty, and Cannot do at all

Source: Washington Group on Disability Statistics, "The Washington Group Short Set on Functioning (WG-SS)," October 11, 2022, https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/.

About the WFP

The World Food Programme provides humanitarian aid through various programs to people in more than 120 countries around the world, primarily during emergencies, such as armed conflicts, natural disasters, and other types of crises. The most common WFP programs provide emergency food or cash transfers to people in urgent need of aid. Food include rice, oil, salt, flour, lentils, and other nutritious foods; cash transfers include cash-in-envelope, mobile money, and vouchers. The WFP sets up distribution sites in central areas, and beneficiaries must come to these sites with identification documents to receive the provisions.

Data on beneficiaries are collected through household surveys during targeting (before they receive benefits) and after distribution of assistance. In the targeting phase, surveys are used to help determine the household's level of vulnerability; having a member of the household with a disability is one criteria of vulnerability. Meeting or surpassing the vulnerability threshold qualifies the household for WFP benefits. After distribution of assistance, surveys are centered around understanding household food security and other outcomes related to the assistance the WFP provides.

The WFP collects household data primarily to quantify a household's vulnerability to food insecurity for beneficiary selection and to understand outcomes related to WFP programs across different types of households, such as households that include a person with a disability, female-headed households, heads-of-households with various levels of education, or large households with many dependents.[5] By contrast, governments use these kinds of data for wide and varied purposes, such as generating population counts; allocating government resources and funding; and developing an array of government policies and programs, or, as Madans, Loeb, and Altman (2011) summarize in their review of the Washington Group, "policy on equalization of opportunities for the population with disabilities."

Feasibility and Purpose of Data Collection

In collaboration with Trinity College Dublin, the WFP field-tested the WG-SS questions in 2020–21 and designed additional guidelines on how to use them to collect disability data (WFP 2022). The guidelines propose five criteria to assess whether the questions should be used: purpose, buy-in, feasibility, quality, and analysis and action. Of these, feasibility and purpose present the greatest barriers to using the questions.

The feasibility criteria require a "yes" response from the WFP staff member to the question: "Is disaggregation feasible in the available timeline and implementation context, using the available resources and modality of data collection?" (WFP 2022, 6). It is important to collect data with a large-enough sample size for disaggregation by disability status to be representative. But when collecting data during emergencies, it is nearly impossible (and infeasible) to acquire representative samples with a household's disability status as a sampling strata in the same way that WFP can acquire representative samples by geographic location, education of the head of household, and sex of the head of household.

The purpose criteria require a "yes" response from the WFP staff member to the question: "Is there a clear and shared understanding of why these data should be collected and how the resulting information can contribute to the programmatic objectives?" (WFP 2022, 6). The latter half of the question is the more important: How can disability data contribute to the WFP programs' objectives? Often, WFP follow-up surveys measure progress on certain outcomes regarding food security, livelihoods, and coping mechanisms. The survey questions include the following:

- How often are major food groups, such as proteins and vegetables, consumed?
- Have children been withdrawn from school because of lack of food or money to buy food?
- Do adults reduce their meal sizes so their children can eat?

The WFP does not have specific program objectives for people with disabilities or households with people with disabilities, apart from enrolling them into its programs for food or cash transfer assistance. This approach differs from other programs with target populations like pregnant and breastfeeding women, where the objective is to help women increase their intake of nutritious foods. Because the WG-SS are often only included in post-distribution surveys, the purposive criteria for inclusion will never be met because they mandate disability data be collected before enrollment. If the WFP is to meet the WG-SS purposive criteria, offices would have to completely change their program design to make disability data key to program objectives.

While there are certainly methodological questions about the use of the WG-SS at the WFP, WFP staff are more often occupied with coordinating real-time, on-the-ground response to reach people in need. Even if the answer to feasibility and purpose questions are "maybe" or a weak "yes," the WG-SS will likely still be included in WFP's household surveys because it is simpler to add the existing question set than to dedicate time to the bigger, strategic questions on data representativity or wider program design.

Using WG-SS Questions in Humanitarian Contexts

The Washington Group discourages adaptations to the WG-SS questions, but these adaptations inevitably occur in organizations like the WFP. While the WFP has organizational guidelines on how survey questions should be asked and how data are collected, cleaned, analyzed, and reported, there is no formal or global mechanism to enforce local use of these guidelines.[6] This variation occurs because each of the 120 WFP country offices (COs) has complete ownership of its survey design, the cost of administering the survey, and the data collected from its surveys. Survey questions (not only the WG-SS) may be adapted to align better with government input or cultural contexts, or to adjust the length of the survey based on time and costs. The independent status of COs is integral to how the WFP operates because the flexibility allows COs to adapt to rapidly changing contexts and the needs of their populations without having to get formal approvals from regional bureaus or headquarters. This flexibility, while beneficial in many respects, also results in numerous adaptations of the WG-SS questions, indicating that the original questions may be unacceptable or suboptimal for use at the field level.

While the WG-SS may be an improvement on the old method of asking about disability, the adaptations of the WG-SS in COs across the globe mean that the disability data collected are not standardized nor comparable, counter to the WG's original intent. Adaptations in and of themselves are not necessarily bad—the WG-SS itself is an adaptation of the longer module—but independent adaptations signal a need for a more efficient way to collect these data. A standardized adaptation of the WG-SS thus needs to be developed—one that preserves the Washington Group's approach but remains relevant and unobtrusive enough for widespread use in humanitarian settings.

To develop a workable version of the WG-SS for WFP COs, it is helpful to understand the various challenges COs face in using the questions, the types of adaptations they make, their reasons for doing so, and the analyses the COs produce.

Why Humanitarian Organizations May Adapt the WG-SS Questions

Individual versus Household Respondents

The six WG-SS questions are designed for individual respondents ages 5 and older, but the WFP's surveys are typically conducted at the household level.[7] Therefore, survey administrators have to decide whether to ask questions only about the head of the household, repeat the questions for each member of the household, or rephrase the questions so one respondent can answer on behalf of the entire household (so-called proxy reporting). The most common choice is to rephrase, from "Do you have difficulty…" to "Does any member of your household have difficulty…".

Length of Time

Asking one or two straightforward questions in a typical household survey can take only a few seconds, whereas asking the WG-SS questions can take about 2 to 3 minutes, which accounts for about 6 to 10 percent of time for an average 30-minute questionnaire. This means that for large surveys—say, a survey with a sample size of 1,000—the WG-SS questions require an additional 33 hours to administer. That is almost a full week of work for an enumerator.

Further, if the WG-SS questions are part of a household roster of questions, they would add anywhere from 6 to 60 more questions for household sizes of 1 to 10. This is extremely lengthy compared with the other questions in the demographic section of a survey, where information such as household size, head of household's highest educational attainment, and number of children under age 5 are typically captured by asking just one question each.

Considering that the WFP surveys are meant to capture outcomes such as food security and livelihoods, unlike a census where demographic data are the focus, front-loading the survey with too many demographic questions would be fatiguing for the respondent and costly to administer.

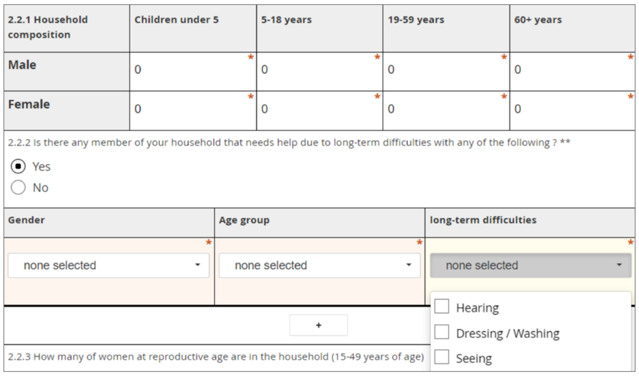
Screening Questions

The Washington Group asserts that asking a screening question before asking the WG-SS questions invalidates the tool, but survey administrators are strongly tempted to do so because it may significantly shorten the time it takes to conduct the survey. For example, a "No" answer to a "Is anyone in your household disabled?" question at the start of the survey would allow the administrator to skip the WG-SS questions entirely. It would also make survey analysis easier, because a Yes/No response allows for summarizing the data on a single variable and avoids having to perform multistep calculations.

World Food Programme staff members took this approach in a survey conducted in five districts of a country in November 2022. They asked, "Is there any member of your household that needs help due to long-term difficulties with any of the following?" then showed options for hearing, dressing, walking, seeing, and communicating. If the response was "Yes," they followed up with questions about the household member's gender and age group and asked the respondent to select the applicable options (**figure 1**).

FIGURE 1

Streamlining WG-SS Questions by Adding a Screening Question



2.2.1 Household composition	Children under 5	5-18 years	19-59 years	60+ years
Male	0 *	0 *	0 *	0 *
Female	0 *	0 *	0 *	0 *

2.2.2 Is there any member of your household that needs help due to long-term difficulties with any of the following ? **

● Yes
○ No

Gender	Age group	long-term difficulties
none selected *	none selected *	none selected *

☐ Hearing
☐ Dressing / Washing
☐ Seeing

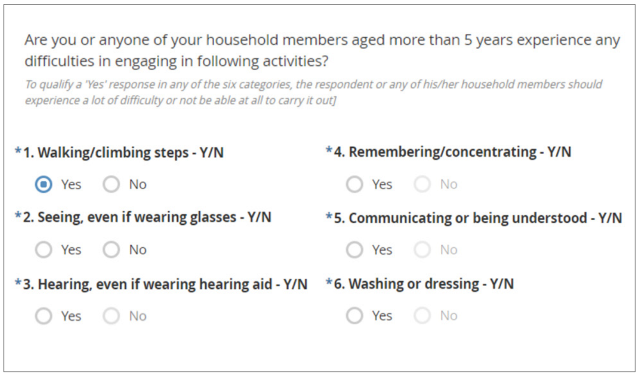
2.2.3 How many of women at reproductive age are in the household (15-49 years of age)

Source: World Food Programme household survey, November 2022.

A second way to shorten the WG-SS questions is to ask about all six functioning areas once using the same phrasing as the WG-SS questions, and allow respondents to answer "Yes" to each area only if their level of difficulty is "A lot of difficulty" or "Cannot do at all." WFP staff members used this approach in a survey of the same five districts of a country in March 2023 (**figure 2**).

FIGURE 2

Streamlining WG-SS Questions without Using a Screening Question



Are you or anyone of your household members aged more than 5 years experience any difficulties in engaging in following activities?

To qualify a 'Yes' response in any of the six categories, the respondent or any of his/her household members should experience a lot of difficulty or not be able at all to carry it out]

*1. Walking/climbing steps - Y/N
● Yes ○ No

*2. Seeing, even if wearing glasses - Y/N
○ Yes ○ No

*3. Hearing, even if wearing hearing aid - Y/N
○ Yes ○ No

*4. Remembering/concentrating - Y/N
○ Yes ○ No

*5. Communicating or being understood - Y/N
○ Yes ○ No

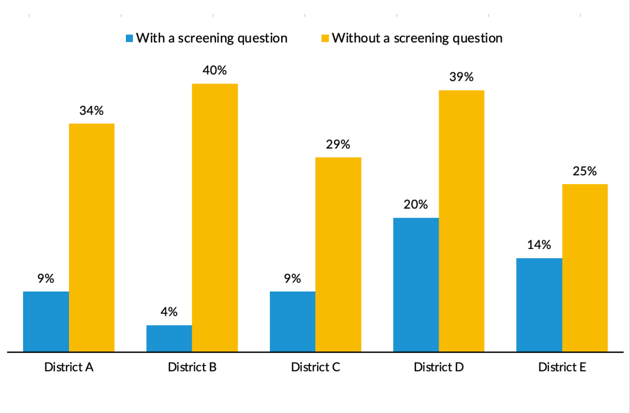
*6. Washing or dressing - Y/N
○ Yes ○ No

Source: World Food Programme household survey, November 2022.

According to the combined results from these surveys (**figure 3**), the share of households reporting that they have someone with a disability is significantly higher across all districts when the survey does not use a screening question. It is unclear, however, whether this difference stems from the way the questions were phrased or the presence of a screening question.[8] Interestingly, an example provided by the Norwegian Refugee Council in a different analysis shows the opposite effect: the share of households self-identifying a person with a disability was 3 percent without a screening question and 18 percent with a screening question (Quigley et al. 2018, 25).

FIGURE 3

Share of Households with a Person with Disability Using the WG-SS Questions, with and without a Screening Question



Source: Excerpted from a post-distribution monitoring report by a World Food Programme (WFP) country office; anonymized for privacy.

Note: Post-distribution monitoring is a standard part of the WFP's monitoring process, where a sample of beneficiaries is surveyed 2–3 weeks after they receive assistance. These reports are then shared internally with other WFP colleagues and sometimes with partners or stakeholders.

Analysis

The six WG-SS questions require calculations that can be confusing for data users, especially when they are repeated for multiple household members because data must be aggregated across different dimensions. First, data are aggregated across the four potential responses to each question to determine if the person has a disability, then across the six questions to see if at least one area of disability is flagged, and then across the household to determine if *any* member of the household is a person with a disability. The WG recommends aggregating the data into two binary variables if necessary: whether the *person* can be classified as a person with a disability and whether the *household* includes a person with a disability.

While analysts clean the data to create the final binary variables, they must also calculate mid-level aggregations, and analysts typically create interim variables to help with the complexity. Data analysts may thus be tempted to produce additional statistics that can answer questions such as the following:

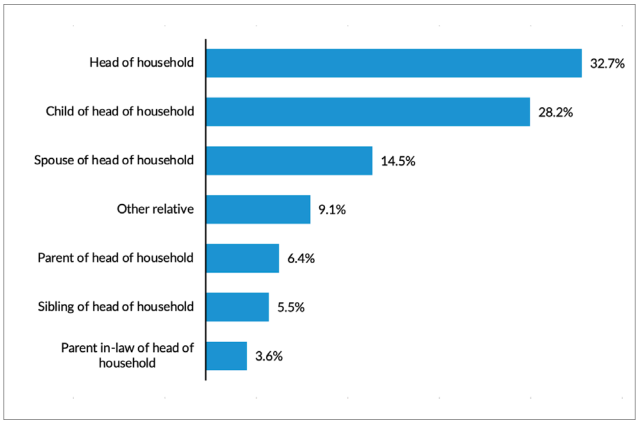
- Which disability/disabilities is/are most prevalent?
- Does a household/head of household have multiple disabilities?
- Does a household have multiple people with multiple disabilities?
- Does a household have multiple people with the same disability?

Because the questions are not intended as an identification tool, the WG does not recommend analyzing the data by specific categories. Nevertheless, WFP technical teams often conduct further analyses, as it seems a waste to collect such granular data only to report on the final aggregated results.

Figures 4 and 5 show additional analyses performed by a WFP CO using an adaptation of the WG-SS data in new ways.

FIGURE 4

Position of People with Disabilities in the Household

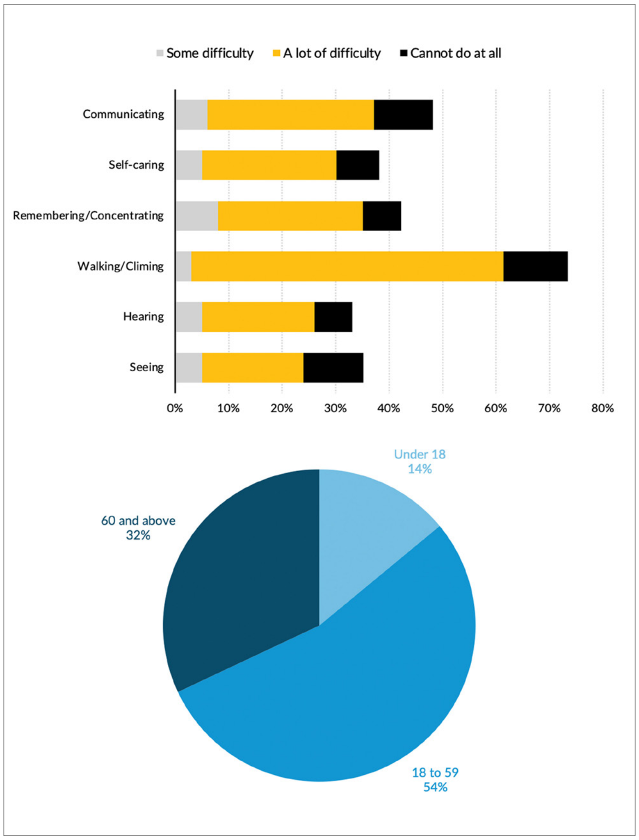


Source: Excerpted from a post-distribution monitoring report by a World Food Programme (WFP) country office.

Note: Post-distribution monitoring is a standard part of the WFP's monitoring process, where a sample of beneficiaries is surveyed 2–3 weeks after they receive assistance. These reports are then shared internally with other WFP colleagues and sometimes with partners or stakeholders.

FIGURE 5

Share of People with Disabilities across Certain Demographic Categories



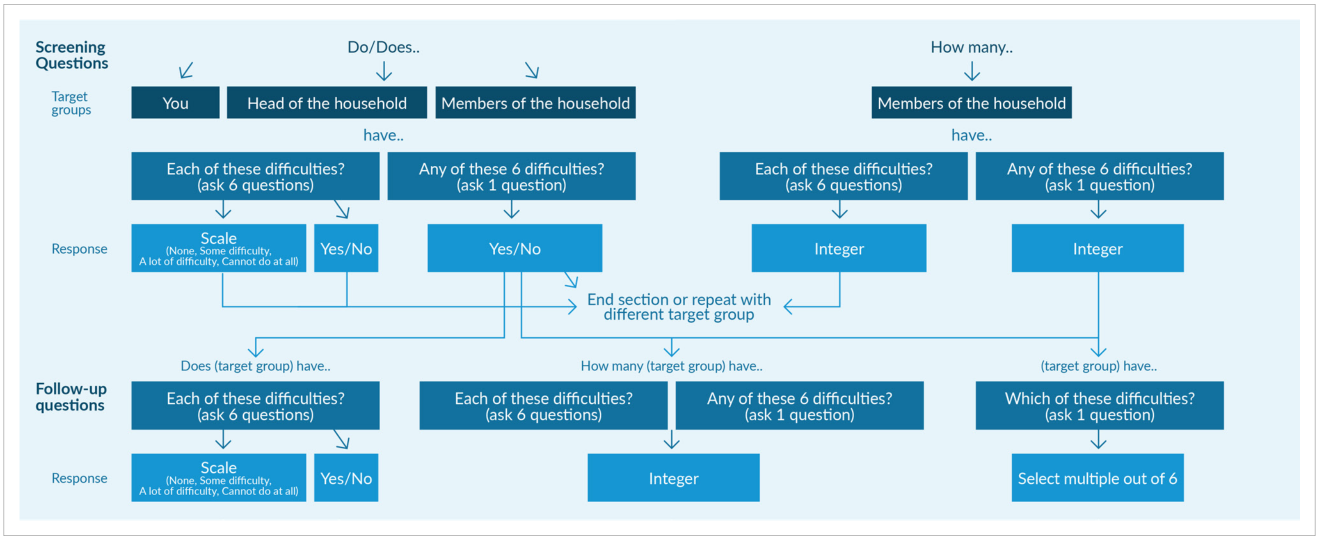
Source: Excerpted from a post-distribution monitoring report by a World Food Programme (WFP) country office.

Note: Post-distribution monitoring is a standard part of the WFP's monitoring process, where a sample of beneficiaries is surveyed 2–3 weeks after they receive assistance. These reports are then shared internally with other WFP colleagues and sometimes with partners or stakeholders.

The WG-SS questions are adapted in ways that attempt to solve some of or all of the challenges country offices face. **Figure 6** shows a flowchart of the different ways WFP COs adapt the WG-SS questions. For example, some adaptations start with "Do you/head of household/member of household" then proceed to ask either one question or all six questions. Others choose to receive responses as a scale, a Yes/No, or a number. The extent of the adaptations and mix-and-match examples demonstrate a need for WG-SS questions that make sense for the WFP and other humanitarian organizations.

FIGURE 6

Various Adaptations of WG-SS Questions by World Food Programme Country Offices



Source: Created by authors.

Recommendations for Data Collection on Disability in Humanitarian Contexts

There is no doubt that the WG-SS questions have caused a paradigm shift in the way all actors—government, humanitarian, or other organizations—think about data collection among people with disabilities (Boggs et al. 2022; Mojtahedi, Mikkola, and Saxén 2023; World Bank 2018). But there is still more work to be done on improving the questions so they can be applied to non-census-based data collection exercises, aggregated to the household level, and used in different humanitarian contexts.

The following are ways the WFP can make disability inclusion more mainstream and better enable accurate household-level data collection.

Create Disability-Sensitive Programs

Designing disability-sensitive programs in humanitarian contexts with reliable and accurate data would be a huge step toward helping people with disabilities. Using disability data in the beneficiary selection process is already well-established in the WFP, but this not sufficient to achieve full inclusion (WFP 2021). Currently, the WFP's programs disaggregate food security indicators by household disability status, which shows that households with a person with a disability have lower levels of food security than other households. Although this disparity demonstrates that the presence of a person with a disability in a household is a determinant of vulnerability—and eligible to receive WFP assistance—few practical steps can be taken after that initial assessment because of a lack of more specific information.

With better, more nuanced, and more reliable data that enable detailed analysis, the WFP could design programs targeting people with disabilities to provide specialized food (e.g., softer foods that are easier to swallow), information in multiple accessible formats (e.g., written, audio, large print), assistance with transportation to and from distribution sites, or even cash transfers to offset the costs of hiring specialized vehicles (WFP 2021, 9). Such measures can clearly help deliver assistance to vulnerable households in a dignified manner and move toward creating truly inclusive programs.

Improve the WG-SS Questions for Use with Households in Humanitarian Contexts

In their recommended format, the WG-SS questions are not suitable for use in humanitarian settings and contexts; many WFP COs informally adapt the questions to address cultural, regional, and situation-specific issues. A different set of standardized questions that are more user-friendly at the field level is needed to collect disability data in practice, not just in theory.

Some minimum criteria for the questions include brevity, collection at the household level, adaptability to evidence or program needs, and ease of analysis. For example, the six WG-SS questions could be consolidated into one question—"How many members in your household have a lot of difficulty in any of the following areas?"—with the response being the number of household members (**box 3**). This change would enable analysis at the binary level (e.g., "Does this household contain any people with disabilities?") as well as on a scale (e.g., "Are households with multiple people with disabilities more vulnerable to food insecurity?").

In addition, this simplifies the responses, instead of having "A lot of difficulty" or "Cannot do at all" representing two answer options that would qualify the person as having that disability. Intuitively, the respondent should understand that having a lot of difficulty in an area would still accurately describe someone who cannot do it at all. Simplifying the text describing each of the six functioning areas can make this single-question adaptation even easier to ask. The original wording may add confusion when translating into local languages, as many WFP beneficiaries may not know or have access to glasses or hearing aids. Wording that is removed can be added to interviewer trainings, or as a note in the questionnaire, so the interviewer can clarify if asked.

If it is important to know the disability status of the head of the household apart from other members of the household, the single-question adaptation can be split in two: keeping the question in box 3 on number of household members, then asking if the head of household is one of the members counted (**box 4**).

These adaptations are suggestions and still need to be field tested. Nevertheless, they may offer a good starting point toward a standardized and consistent adaptation of the WG-SS that minimizes the burden of data collection in humanitarian settings while preserving the spirit and approach of the WG-SS.

The original WG-SS questions about individual functioning areas should be asked only if the aim is to collect data that will inform program decisions, looking back at the purposive criteria mentioned earlier. For example, separate data on beneficiaries with difficulty walking should be collected only if the WFP or its cooperating partners intend to provide additional transportation assistance at distribution sites and need to know the areas with a higher prevalence of people with disabilities needing this help. Similarly, the WFP can hire sign language interpreters or assistants specifically to help people who are hard of hearing or blind.

Unless more sophisticated disability-inclusive programming becomes a part of WFP's mainstream programs, time and resources should not be spent collecting detailed data on disability per the WG-SS; a simplified adaptation like the ones recommended above should be used instead.

BOX 3: CONSOLIDATING THE SIX WG-SS QUESTIONS INTO ONE QUESTION

How many members in your household have a lot of difficulty in any of the following areas?

- seeing
- hearing
- walking or climbing steps
- remembering
- washing all over or dressing
- communicating

Response types: Integer; number of household members

BOX 4: CONSOLIDATING THE SIX WG-SS QUESTIONS INTO TWO QUESTIONS

1. How many members in your household have a lot of difficulty in any of the following areas?

- seeing
- hearing
- walking or climbing steps
- remembering
- washing all over or dressing
- communicating

Response types: Integer; number of household members

2. Is the head of the household one of those members?

Response type: Yes/No

Conclusion

Since the WG-SS was created over two decades ago, the collection and use of data in humanitarian settings have grown exponentially. But the WG-SS is poorly suited for use in these contexts. The length of time required to administer the WG-SS poses the greatest challenge, especially in household surveys where collection of disability data is not the primary objective. Additionally, the complexity of the WG-SS data collected and the simplistic binary results the WG recommend create a disconnect, either by tempting detailed analyses or not offering valuable enough insight relative to its cost.

Thus, a simplified version of the WG-SS is needed to meet the needs of humanitarian organizations like the WFP. Once better data on disability are collected, humanitarian interventions, such as food and nutrition security assistance programs by the WFP and other stakeholders, will be better informed to increase inclusion and accessibility for people with disabilities.

NOTES

1. The agenda and other seminar materials for the International Seminar on Measurement of Disability, New York, June 4–6, 2001, are available at the United Nations Statistics Division, accessed July 18, 2024, https://unstats.un.org/unsd/disability/Seminar%202001.html.
2. Washington Group on Disability Statistics, “Question Sets,” accessed January 25, 2024, https://www.washingtongroup-disability.com/question-sets.
3. Washington Group on Disability Statistics, “The Washington Group Short Set on Functioning (WG-SS),” October 11, 2022, https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/.
4. The United Nations World Food Programme is “the world’s largest humanitarian organization saving lives in emergencies and using food assistance to build a pathway to peace, stability and prosperity, for people recovering from conflict, disasters and the impact of climate change.” World Food Programme, “Who We Are,” accessed July 18, 2024, https://www.wfp.org/who-we-are.
5. Food security analyses performed by WFP and partners in different contexts seem to confirm the correlation between being disabled (or having someone with a disability in the household) and vulnerability to food insecurity. However, researchers, enumerators, and respondents have varied interpretations of disability.
6. There are formal methods for administering corporate outcome indicators correctly. However, the WG-SS data are considered demographic indicators, so they fall outside the corporate reporting requirements.
7. For many WFP surveys, the average household size ranges from 5 to more than 10.
8. Figures 3, 4, and 5 are excerpted from post-distribution monitoring reports by WFP country offices. There are issues with the validity of analyses in such reports, but they are presented in this essay to illustrate the ways country offices use the WG-SS data.

REFERENCES

Boggs, Dorothy, Hannah Kuper, Islay Mactaggart, Tess Bright, GVS Murthy, Abba Hydara, Ian McCormick, et al. 2022. “Exploring the Use of Washington Group Questions to Identify People with Clinical Impairments Who Need Services including Assistive Products: Results from Five Population-Based Surveys.” *International Journal of Environmental Research and Public Health* 19 (7): 4304. https://doi.org/10.3390/ijerph19074304.

Landes, Scott D., Bonnielin K. Swenor, and Nastassia Vaitsiakhovich. 2024. “Counting Disability in the National Health Interview Survey and Its Consequence: Comparing the American Community Survey to the Washington Group Disability Measures.” *Disability and Health Journal* 17 (2): 101553. https://doi.org/10.1016/j.dhjo.2023.101553.

Madans, Jennifer H., Mitchell E. Loeb, and Barbara M. Altman. 2011. “Measuring Disability and Monitoring the UN Convention on the Rights of Persons with Disabilities: The Work of the Washington Group on Disability Statistics.” *BMC Public Health* 11 (Suppl 4). https://doi.org/10.1186/1471-2458-11-S4-S4.

Mojtahedi, Mina C., Riikka Mikkola, and Niklas Saxén. 2023. “Advancing Rights and Inclusion of Persons with Disabilities in Humanitarian Action: A Donor Perspective.” *International Review of the Red Cross* 105 (922): 416–33. https://doi.org/10.1017/S1816383122000868.

Quigley, Nolan, Emma Bird, Kate Turner, Gemma Cook, and Pauline Thivillier. 2018. *Disability Data Collection: A Summary Review of the Use of the Washington Group Questions by Development and Humanitarian Actors*. London: Leonard Cheshire and Humanity & Inclusion.

WFP (World Food Programme). 2021. *Mainstreaming of Disability Disaggregation: A Phased Approach*; Technical Note. Rome: WFP.

———. 2022. *Disability Data: An Evidence-Informed Approach to the Use of Disability Disaggregated Data in WFP Programming*. Rome: WFP.

WG (Washington Group on Disability Statistics). 2020. *The Washington Group Primer: An Introduction to the Washington Group on Disability Statistics Question Sets*. Hyattsville, MD: Washington Group.

World Bank. 2018. "*World Bank Group Commitments on Disability-Inclusive Development*." Washington, DC: World Bank.

Chapter Three

Advocating for Greater Data Inclusion for Puerto Rico

CHAPTER THREE

Advocating for Greater Data Inclusion for Puerto Rico

JULIA BIEDRY GONZALEZ

Gathering accurate, complete, and comprehensive data is a cornerstone of effective governance and policymaking because data provide the foundation for critical decisionmaking. While data are supposed to represent quantitative facts, many qualitative factors can affect how those "facts" are formed, including how data are collected and how results are communicated. When data methodologies are unequally applied, inequities in policymaking can be amplified.

In Puerto Rico, an archipelago commonwealth of the United States, data collection is irregular and disparate across several dimensions compared with data collection in the mainland US. Puerto Rico's nuanced historical, political, and cultural relationship with the mainland United States has long affected how its data are collected for annual data products, including the US Census, which in turn affects how the data are analyzed and presented. These inconsistencies not only hinder Puerto Rico's development and affect the aid and funding it receives from the federal government but call into question how equity and representation are applied to Puerto Rico and the rest of the US island territories.

To understand how to better include Puerto Rico in key US data products that drive policy and funding decisions, this essay explores Puerto Rico's unique status and key gaps in federal data collection.

History of Puerto Rico's Status in the US

Puerto Rico has had a complicated and ever-changing relationship with the US government for more than a century. The archipelago first became a US territory under the Treaty of Paris after the Spanish-American War of 1898 (Cheatham and Roy 2022). In 1917, Puerto Ricans became US citizens via the Jones-Shafroth Act but were not granted the ability to fully participate in federal elections.[1] In 1952, the US government authorized Puerto Rico to ratify its own constitution and the archipelago became a commonwealth of the United States, officially known as the *Estado Libre Asociado* (Free Associated State).[2] This status granted a degree of self-governance to the archipelago, although the US maintained ultimate control.

Positionality Statement

Julia Biedry Gonzalez is a data professional who grew up in New York. Her mother is Puerto Rican with family roots in Lares, Puerto Rico. Biedry Gonzalez has worked extensively in public-sector and nonprofit analytics consulting. This essay is born from her own efforts to include Puerto Rico and the US territories in her analyses and her realizations about the structural hurdles involved in doing so.

Since then, Puerto Rico's relationship to the US remains an active subject of debate with opposing advocates calling for statehood, for independence, or for maintaining the commonwealth arrangement.[3] Puerto Ricans' lack of full representation in Congress and inability to vote in presidential elections have presented wide-ranging issues for the territory—issues mirrored in its history of federal data collection.

Conditions in present-day Puerto Rico reflect the effects of its complicated history and US policymaking that has never fully accounted for the needs of Puerto Ricans. The Merchant Marine Act of 1920, for instance, requires that goods be carried between US ports on US ships. This limitation on trade costs Puerto Rico an estimated $1.4 billion annually and contributes to the archipelago's high cost of living and depressed development (Hillberry and Jimenez 2024). Puerto Rico has a higher poverty rate than any US state, yet its residents are excluded from many federal benefits, while the number of available jobs continues to decrease (Vargas-Ramos et al. 2023).[4] Since 2014, Puerto Rico has been subject to a fiscal control board, put in place because of Puerto Rico's deteriorating financial situation, which is exacerbated by the lack of reliable economic statistics to drive the local government's decisionmaking.[5] These unique challenges continue to highlight the need for accurate statistics—to measure population change, to appropriately fund welfare programs, to assess the effects of legislation, and to fully represent Puerto Ricans in government policy as what they are: US citizens.

Inconsistent Data Collection in Puerto Rico

Although federal data collection includes Puerto Rico more often than the other island territories (Guam, the Northern Mariana Islands, US Virgin Islands, and American Samoa), data collection processes in Puerto Rico diverge significantly from processes in the 50 states. Data about Puerto Rico are collected differently, provided separately, or left out at every step of the analytics process—from collection to presentation. When data are provided separately or included in some products but not others, data analysts and scientists working in the federal government and nongovernmental organizations have more difficulty harmonizing data to better understand and influence US policy or funding decisions. They are forced to take additional steps to incorporate Puerto Rico (as well as the other island territories) into their analyses of US data.

Furthermore, incremental changes to data collection in Puerto Rico have made time-series comparisons within Puerto Rico and with the 50 states difficult. When analysts use large-scale datasets or rely on automation, it is even easier to overlook the lack of Puerto Rican data; information may exist in only one dataset or at different levels of aggregation, often resulting in the exclusion of Puerto Rico data altogether.

The lack of regulation, standardization, and harmonization of datasets creates a hurdle for representing Puerto Rico equitably in US policymaking and budgeting decisions. This issue is compounded by Puerto Rico's status in the US and its lack of representation in Congress. Gaps in data create gaps in equity. Not collecting the data, not collecting data often enough, or not collecting granular data obscures the realities and hardships of Puerto Ricans, all of whom are US citizens and a majority of whom identify as a race other than white (Figueroa-Lazu, Hinojosa, and Bonilla 2022).

Exclusion from Federal Data Collection and Analysis Products

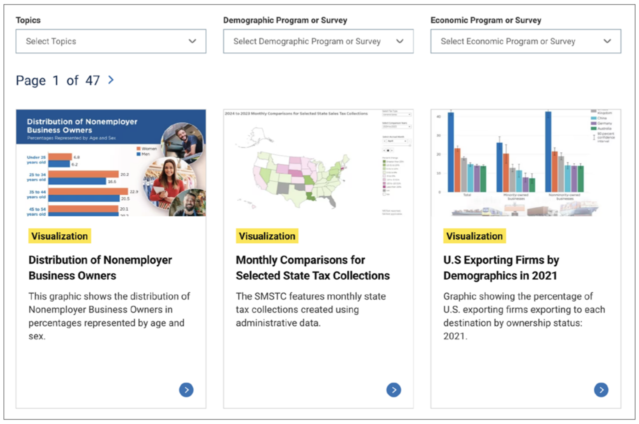
Among the many annual federal surveys, Puerto Rico is included in the Census Bureau's decennial Census, the American Community Survey (through a separate survey instrument), and population estimates. Other island territories are excluded from all but the decennial Census, and all island territories including Puerto Rico are excluded from the Current Population Survey.[6] Puerto Rico is not tracked as part of the US gross domestic product or even in the Census Bureau's count of the US population.[7]

Puerto Rico was first included in the US Census in 1910, seven years before Puerto Ricans were officially designated as US citizens. However, Puerto Rican residents only began receiving the Census form like residents of US states in 2000.[8] Puerto Rican residents are included in the American Community Survey through a separate survey called the Puerto Rico Community Survey, which began in 2005 (US Census Bureau 2020). The two surveys use different data collection methodologies. For the Puerto Rico Community Survey, data are only collected by mail or in person, not via the internet; estimates are not controlled for race or Hispanic origin; subcounty information is not available; and housing unit controls are not available (US Census Bureau, n.d.). Furthermore, the Census collects detailed demographic data on age and sex but not race or ethnicity in Puerto Rico.[9]

Other major Census products exclude Puerto Rico entirely, such as the Current Population Survey,[10] which the Bureau of Labor Statistics uses to produce important measures of the economy like the poverty rate, and the Demographic Analysis, which is used to calculate net coverage error of the Census and is informed by current and historical vital records, international migration, and Medicare reports (US Census Bureau 2022). Data on Puerto Rico also are often left out of downstream analyses presented by the Census Bureau, the Centers for Disease Control and Prevention, and other government agencies and partners. As of June 2024, none of the 17 US maps on the first page of the Census Infographics and Visualization library include Puerto Rico (**figure 1**). Even when data on Puerto Rico are collected, they are often presented separately, such as in the Centers for Disease Control and Prevention's Social Vulnerability Index (first published in 2000, with Puerto Rico added in 2014; see ATSDR 2022) and the Census Community Resilience Estimates.

FIGURE 1

Maps of the United States on the Census Bureau's Infographics and Visualization Web



Distribution of Nonemployer Business Owners

This graphic shows the distribution of Nonemployer Business Owners in percentages represented by age and sex.

Monthly Comparisons for Selected State Tax Collections

The SMSTC features monthly state tax collections created using administrative data.

U.S Exporting Firms by Demographics in 2021

Graphic showing the percentage of U.S. exporting firms exporting to each destination by ownership status: 2021.

Source: US Census Bureau, "Infographics & Visualizations," accessed January 9, 2024, https://www.census.gov/library/visualizations.html.

The Community Resilience Estimates analysis perfectly demonstrates how equity problems can be obscured when comprehensive data are not collected or presented. Although the Census has been creating Community Resilience Estimates for the 50 states for five years, it first produced a similar estimate for Puerto Rico in 2023. The results were astonishing: 46.1 percent of Puerto Ricans suffer from three or more risk factors, such as poverty, lack of a high school degree, advanced age (65 or older), or lack of health care coverage. That number more than doubles the rate in the 50 states (21.6 percent), and every municipio (county-level equivalent) in Puerto Rico had higher rates of people with three or more risk factors than the average in the mainland US.[11]

These gaps highlight the need to include Puerto Rico in federal data products.[12] Boricua organizations in Puerto Rico and the mainland—including the Puerto Rico Institute of Statistics (PRIS), Instituto del Desarrollo de la Juventud, the Centro para una Nueva Economia, and the Center for Puerto Rican Studies at Hunter College—have been aware of these gaps for years and have worked to create accurate, reliable data and analytics for Puerto Rico.

This essay is by no means a comprehensive catalog of the data gaps faced by Puerto Rico. More research is needed to quantify the true extent of the gaps at every step of the analytics process *and* to enumerate all federal data and analytics products that exclude Puerto Rico and the other island territories.

The Debt Crisis, Medicaid, and Data

A lack of accurate statistics is deeply intertwined with challenges in access to benefits and, by extension, the debt crisis in Puerto Rico. While the Center for Medicaid and CHIP Services collects data on Puerto Rico as part of its Transformed Medicaid Statistical Information System analytic files, it excludes Puerto Rico from its monthly Performance Indicators data collection, which includes data on Medicaid and CHIP enrollment activity and program performance.[13] This exclusion limits the federal government's access to periodic, accurate, up-to-date data on Puerto Rico and has downstream effects for the territory. For 2021, Puerto Rico scored "high concern," "unusable," or "unclassified" on nearly half the top-line data quality indicators (57 of 119) presented in the Center for Medicaid and CHIP Services' Data Quality Atlas because of the lack of comparison data from the performance indicators and other sources.[14]

This lack of up-to-date quality data is notable because of the crisis facing Puerto Rico's health system. Perreira and colleagues (2017) found that the island was reliant on Medicaid, and its health care system has been crippled by many issues, including brain drain to the mainland, economic instability, high cost of living, and the rapidly aging population. Despite its high need, Puerto Rico receives a capped amount of Medicaid assistance, in contrast to the uncapped scaling system for the 50 states (Stolyar, Orgera, and Rudowitz 2021). In 2021, Representative Nydia Velázquez (D-NY) and Senator Bernie Sanders (I-VT) introduced the Territorial Equity Act to address the unequal Medicaid assistance and fund the Supplemental Nutritional Assistance Program in the US territories.[15] The bill, however, has yet to be brought up for a vote. Puerto Rico's exclusion from uncapped Medicaid and the Supplemental Nutritional Assistance Program leads to higher expenditures by the local government, which is cited as a major driver of the territory's ongoing debt crisis.[16]

Beyond health care and nutrition benefits, unreliable economic data have played into Puerto Rico's debt crisis in other ways. Lack of data skills and funding, overestimated inflation and tax revenues, inaccurately measured tourism activity, and insufficient awareness of population loss to the mainland have all been cited as additional reasons for the crisis, all caused by an inadequate emphasis on collecting reliable data (see, for example, Langkjaer-Bain 2018).[17]

Hurricane Maria and the Impact of Data

Issues with data collection in Puerto Rico have been exacerbated and underscored by recent disasters, most notably Hurricane Maria, which destroyed homes and caused power outages that stretched for as long as six months in 2017.[18] The human toll was even more devastating: an estimated 4,645 people lost their lives.[19] However, the official government death count immediately after the storm was just 64 people, which was revised a year later to 2,975 following an independent study.[20] The gross initial undercount spotlights the lack of reputable data collection in Puerto Rico.

Before Hurricane Maria hit, the Puerto Rican government was attempting to disband the Puerto Rico Institute of Statistics, an independent public research institute, because of federal funding shortfalls. The government instead directed its Department of Public Safety to lead the information gathering on deaths. Mario Marazzi, executive director of the institute, stated that he believed the death toll in the three months after Hurricane Maria topped 1,000.[21] Marazzi also penned an op-ed highlighting Puerto Rico's exclusion from major US federal data projects and the need for an "inordinate amount of data transparency and validation, if Puerto Rico is going to be able to come out of this mess."[22] The controversy following the initial undercount by the Puerto Rican government brought widespread attention to the territory's systemic data reporting deficiencies.

Hurricane Maria affected the accurate and timely collection of Census products as well. In 2022, the US Government Accountability Office found that major crises like hurricanes (as well as the COVID-19 pandemic) led to widespread data collection challenges in the territories.[23] The Census Bureau, for example, reported that Hurricane Maria affected Puerto Rico Community Survey data collection in 2017 (Jensen et al. 2020). The downstream effects of that inconsistent federal data collection are still felt in Puerto Rico today. A headline-generating study recently published in *Science Advances* about excess mortality following tropical cyclones in the US completely excluded Hurricane Maria from its analysis. The authors acknowledged this exclusion: “This study focused on the continental United States (CONUS), although devastating effects of tropical cyclones, such as Hurricane Maria in Puerto Rico, have also been recorded. This was in part because currently publicly available, curated tropical cyclone, temperature, and precipitation data only cover the CONUS, with SVI [Social Vulnerability Index] also covering CONUS and Puerto Rico; further work should increase the scope to other US territories and other parts of the world” (Parks et al. 2023, 6 of 12).

Since Hurricane Maria, Puerto Rico has continued to face natural disasters; its location, climate, and geography leave it more exposed to natural disasters than many other parts of the US and, indeed, the world. The 2021 Global Climate Risk index ranked Puerto Rico, alongside Haiti and Myanmar, as the most affected by weather-related loss events from 2000 to 2019 (Eckstein, Künzel, and Schäfer 2021). Disasters have and will continue to affect Puerto Rico, regardless of whether accurate data are captured about the true extent of damage to its citizens, economy, and way of life. Since Hurricane Maria, Puerto Rico has weathered additional hurricanes, earthquakes, and tropical storms. Puerto Rico’s unique vulnerability to natural disasters only deepens the need for accurate statistics, reporting, and, subsequently, policymaking and aid to better protect Puerto Ricans.

Greater Inclusion for Puerto Rico in Data, Analytics, and Data Visualization

Puerto Rico has been included in major federal statistical projects more often than the other island territories because of legislation and executive action. Notably, in 1992, President George H. W. Bush issued an executive action directing “all federal departments, agencies, and officials” to “treat Puerto Rico administratively as if it were a state” in major federal data collection practices (US Department of Justice 2007, 21). A congressional hearing in 2008 highlighted the data issues faced by Puerto Rico and the territories, specifically in regard to Census data.[24] In 2022, Representative Raúl Grijalva (D-AZ) directed the Government Accountability Office to comprehensively review data collection gaps in the territories and their downstream effects on funding. In 2023, Grijalva, in concert with the representatives of Puerto Rico and the island territories, introduced the Territories Statistics Collection Equity Act, which would direct the Interagency Council on Statistical Policy to “collect and publish statistics regarding the US territories in the same manner as states.”[25]

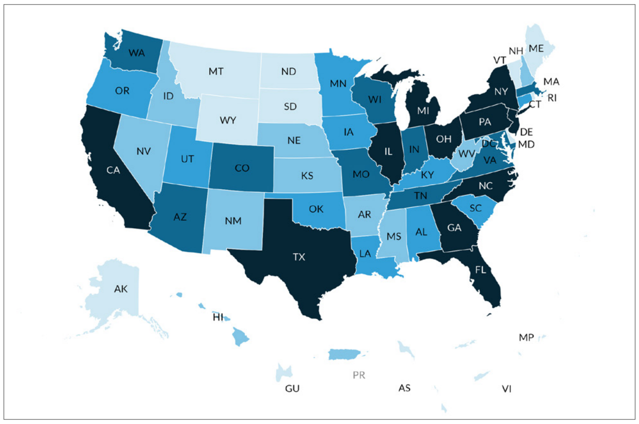
Although this essay focuses on Puerto Rico, many of the same and even deeper data inequities also exist for Guam, the Northern Mariana Islands, the US Virgin Islands, and American Samoa. The Territories Statistics Collection Equity Act is designed to address issues in all five permanently inhabited territories, and policymakers should consider passing the legislation to create more equitable data collection and information sharing practices around US territories. Researchers should also take necessary steps to harmonize and integrate Puerto Rico data into their analyses when those data are available. When researchers collect their own data to design US-based research studies, they should strive to include Puerto Rico. They should also consult with local organizations to understand the best ways to collect these data and translate materials into appropriate languages as needed to reach Puerto Rican residents.

Data analysts could consider several design choices when creating visualizations to encourage greater inclusion of Puerto Rico in datasets and analyses. For map visualizations, data practitioners can use a similar strategy that which is commonly used for Alaska and Hawaii: picturing them alongside the mainland, though often not to scale (**figure 2**). In choropleth maps—maps that use differences in colors or shading to indicate geographic values—the scale often affects Hawaii and makes the color legend or data-related shading unreadable; the same is true for smaller mainland states like Rhode Island. Puerto Rico is larger than Delaware and Rhode Island in land mass and has a population greater than that of 21 states.[26] Prioritizing land size versus population size is a common issue in map-based visualizations, and it is often debated around ballot collection and district mapping during US presidential races (Schwabish 2021).

Data practitioners could also represent the US as a hex map or tile grid map, where each state or territory is given equal representation (**figure 3**).

FIGURE 2

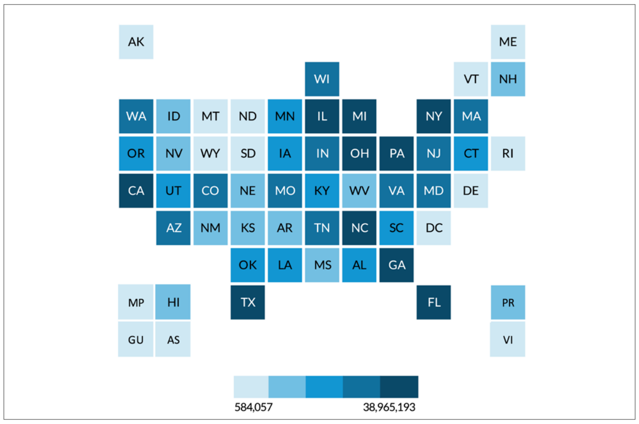
Total 2023 Population for All US States and Territories as a Choropleth Map



Source: Data for US states and Puerto Rico are from US Census Bureau, "State Population Totals and Components of Change: 2020–2023," last revised December 18, 2023, https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html. Data for American Samoa, Guam, the Northern Mariana Islands, and the US Virgin Islands are from the US Census Bureau, International Database, accessed December 1, 2023, https://bit.ly/4dAoN7r.

FIGURE 3

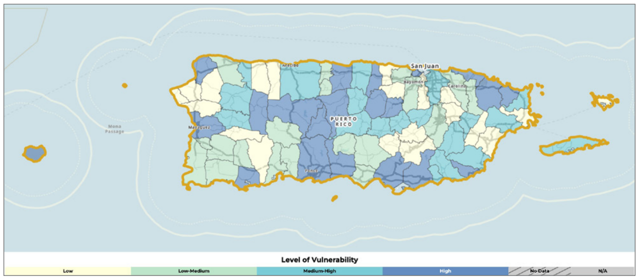
Total 2023 Population for All US States and Territories as a Tile Grid Map



Source: Data for US states and Puerto Rico are from the US Census Bureau, "State Population Totals and Components of Change: 2020–2023," last revised December 18, 2023, https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html. Data for American Samoa, Guam, the Northern Mariana Islands, and the US Virgin Islands are from the US Census Bureau, International Database, accessed December 1, 2023, https://bit.ly/4dAoN7r.

FIGURE 4

CDC/ATSDR Social Vulnerability Index for Puerto Rico, 2022

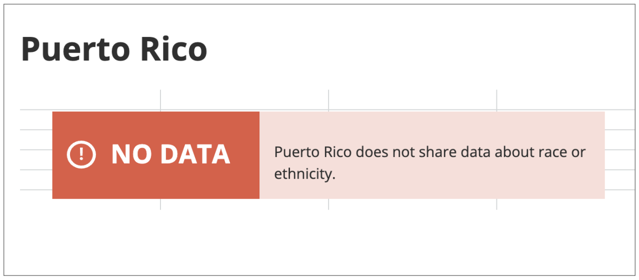


Source: "CDC/ATSDR Social Vulnerability Index," Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry, last reviewed December 1, 2022, https://www.atsdr.cdc.gov/placeandhealth/svi/interactive_map.html.

When substate data are made available, the equivalent should also be shown for Puerto Rico. Municipios are often used as the equivalent to US counties, and their data can allow analyses to drill deeper into local statistics. As is the case with states, conditions and experiences vary across Puerto Rico, and disaggregated local data are needed to understand how communities' needs differ. The CDC's Social Vulnerability Index, for example, shows municipio-level data of various drivers of vulnerability (**figure 4**).

FIGURE 5

Visualization for When Data Are Missing



Puerto Rico

NO DATA

Puerto Rico does not share data about race or ethnicity.

Source: The COVID Tracking Project at *The Atlantic*, "Racial Data Dashboard," accessed January 10, 2024, https://covidtracking.com/race/dashboard#state-pr. The image is licensed under a CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

When visualizing US-based analyses, designers can highlight the gaps in the dataset on Puerto Rico by including footnotes and different colors. When possible, researchers can also highlight the assumptions made and the limitations of the dataset. *The Atlantic*'s COVID Tracking Project's "Racial Data Dashboard" offers a good example. When data are missing for a state or territory, the dashboard provides a clear warning and an explanation (**figure 5**). The website also provides a detailed breakdown of the potential data quality issues within each state's or territory's methodology.

Conclusion

Puerto Rico has long faced gaps in representation and in equity as a territory of the United States, and that complicated relationship is reflected in federal data collection policy that excludes Puerto Rico (and American Samoa, Guam, the Northern Mariana Islands, and the US Virgin Islands) at every step of the analytics process. These gaps exist in the federal data used to make important policy decisions, such as those collected by the Census Bureau and the Center for Medicaid and CHIP Services. Shocking disparities exist even in federal analyses that include Puerto Rico, which further underlines the necessity of illuminating the realities Puerto Ricans face with accurate, up-to-date, and comprehensive data collection and presentation.

It is the responsibility of federal statistical agencies to report equitably on each and every resident of the United States, regardless of where they live. It is the responsibility of federal policymakers to create policies that account for differences in lived experience, which can only be done with accurate data. And it is the responsibility of researchers to include the whole United States in their analyses, not just the 50 states. At every step of the analytics process, scientists, analysts, and policymakers should strive to ensure greater data representation for the people of Puerto Rico and the rest of the island territories.

NOTES

1. Puerto Rico residents are eligible to vote only in the US primaries, not in general elections. "1917: Jones-Shafroth Act," Library of Congress, accessed January 8, 2024, https://guides.loc.gov/latinx-civil-rights/jones-shafroth-act.
2. Magaly Rivera, "Estado Libre Associate (ELA)," Welcome to Puerto Rico, accessed February 20, 2024, https://welcome.topuertorico.org/reference/ela.shtml.
3. Norberto Muniz, "The Unfinished Debate on the Political Status of Puerto Rico and Its Relationship with the United States," Georgetown Law, SALPAL (Georgetown Project on State and Local Government Policy and Law), accessed September 3, 2024, https://www.law.georgetown.edu/salpal/the-unfinished-debate-on-the-political-status-of-puerto-rico-and-its-relationship-with-the-united-states/.
4. Andrew Van Dam, "People Are Fleeing Puerto Rico, Guam and Every Other US Territory. What Gives?" *Washington Post*, September 23, 2022, https://www.washingtonpost.com/business/2022/09/23/american-territories-population-loss/.
5. "About Us," Financial Oversight and Management Board for Puerto Rico, accessed August 28, 2024, https://oversightboard.pr.gov/about-us/.
6. US Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey: Concepts and Definitions," last modified November 28, 2023, https://www.bls.gov/cps/definitions.htm#geo.
7. Jay L. Zagorsky, "Why Puerto Rico 'Doesn't Count' to the US Government," The Conversation (US), October 25, 2017, https://theconversation.com/why-puerto-rico-doesnt-count-to-the-us-government-86262.
8. "Island Areas," US Census Bureau, accessed January 8, 2024, https://www.census.gov/history/www/programs/geography/island_areas.html.
9. US Census Bureau, "Puerto Rico Commonwealth Population by Characteristics: 2020–2023," last revised June 25, 2024, https://www.census.gov/data/tables/time-series/demo/popest/2020s-detail-puerto-rico.html.
10. US Bureau of Labor Statistics, "Labor Force Statistics from the Current Population Survey: Concepts and Definitions (CPS)," last modified November 28, 2023, https://www.bls.gov/cps/definitions.htm#geo.
11. Benjamin Gurrentz, James Scurry, and Isabelle R. Notter, "New Community Resilience Estimates Show Almost Half of Puerto Rico Population Had Three or More Risk Factors," US Census Bureau, June 7, 2023, https://www.census.gov/library/stories/2023/06/how-socially-vulnerable-is-puerto-rico-to-disasters.html.
12. Letter from Mario Marazzi-Santiago (Executive Director, Instituto de Estadísticas de Puerto Rico) to Rep. Nydia Velázquez, "Subject: The Future of Puerto Rico's Economy," July 18, 2017, https://estadisticas.pr/files/Memoriales/memorial_20170718.pdf.
13. Data Quality Atlas, "Total Medicaid and CHIP Enrollment," Center for Medicaid and CHIP Services, accessed January 9, 2024, https://www.medicaid.gov/dq-atlas/landing/topics/single/table?topic=g1m7&tafVersionId=35; and "Monthly Medicaid & CHIP Application, Eligibility Determination, and Enrollment Reports & Data," Center for Medicaid and CHIP Services, accessed January 9, 2024, https://www.medicaid.gov/medicaid/national-medicaid-chip-program-information/medicaid-chip-enrollment-data/monthly-medicaid-chip-application-eligibility-determination-and-enrollment-reports-data/index.html.
14. Data Quality Atlas, "Puerto Rico," Center for Medicaid and CHIP Services, accessed January 9, 2024, https://www.medicaid.gov/dq-atlas/landing/states/single?state=72&tafVersionId=35.
15. Congresswoman Nydia Velázquez (website), "Rep. Velázquez, Senator Sanders Introduce Territorial Equity Act of 2021, Sweeping Reform for Parity for U.S. Territories," press release, April 20, 2021, https://velazquez.house.gov/media-center/press-releases/rep-velazquez-senator-sanders-introduce-territorial-equity-act-2021.
16. Greg Depersio, "The Origins of the Puerto Rican Debt Crisis," Investopedia, updated January 31, 2022, https://www.investopedia.com/articles/investing/090915/origins-puerto-rican-debt-crisis.asp.
17. See also Arnaldo Cruz, "The Other Hidden Crisis in Puerto Rico: Economic Statistics," The Bond Buyer, July 31, 2017, https://www.bondbuyer.com/opinion/the-other-hidden-crisis-in-puerto-rico-economic-statistics.
18. Denise Lu and Chris Alcantara, "After Hurricane Maria, Puerto Rico Was in the Dark for 181 Days, 6 Hours, and 45 Minutes," *Washington Post*, updated April 4, 2018, https://www.washingtonpost.com/graphics/2017/national/puerto-rico-hurricane-recovery/.
19. Daniel Cusick, "Hurricane Maria Contributed to Nearly 5,000 Deaths, Researchers Say," *Scientific American*, May 30, 2018, https://www.scientificamerican.com/article/hurricane-maria-contributed-to-nearly-5-000-deaths-researchers-say/.
20. Nicole Darrah, "Puerto Rico Governor Raises Hurricane Maria Death Toll from 64 to 2,975," Fox News, August 28, 2018, https://www.foxnews.com/us/puerto-rico-governor-raises-hurricane-maria-death-toll-from-64-to-2975.
21. Eileen Drage O'Reilly, "Death Toll from Hurricane Maria Severely Underreported Due to Data Shortfall," Axios, April 3, 2018, https://www.axios.com/2018/04/03/puerto-rico-hurricane-maria-death-toll-underreported-lack-of-data.
22. Mario Marazzi, "Puerto Rico Institute of Statistics Vastly Improved Transparency, Now the Government Is Shutting It Down," The Hill, March 16, 2018, https://thehill.com/opinion/civil-rights/378804-puerto-rico-institute-of-statistics-vastly-improved-transparency-now-the/.
23. US House Committee on Natural Resources, "Grijalva, Territorial Delegates Introduce Bipartisan Bill to Address Disparities in Data Collection for U.S. Territories," press release, March 7, 2023, https://democrats-naturalresources.house.gov/media/press-releases/grijalva-territorial-delegates-introduce-bipartisan-bill-to-address-disparities-in-data-collection-for-us-territories.

24. *Census Data: Special Issues Related to U.S. Territories, Joint Hearing before the Subcommittee on Information Policy, Census, and National Archives and the Subcommittee on Insular Affairs of the Committee on Natural Resources*, 110th Cong., 2d Sess. (2008), https://www.govinfo.gov/content/pkg/CHRG-110hhrg44911/html/CHRG-110hhrg44911.htm.

25. US House Committee on Natural Resources, "Grijalva, Territorial Delegates Introduce Bipartisan Bill."

26. Ryan Struyk, "More Americans Live in Puerto Rico than 21 States (and Other Things to Know)," CNN, September 26, 2017, https://www.cnn.com/2017/09/26/politics/puerto-rico-state-hurricane-maria/index.html.

REFERENCES

ATSDR (Agency for Toxic Substances and Disease Registry). 2022. "CDC/ATSDR SVI 2020 Documentation." Washington DC: ATSDR.

Cheatham, Amelia, and Diana Roy. 2022. "Puerto Rico: A U.S. Territory in Crisis." New York: Council on Foreign Relations.

Eckstein, David, Vera Künzel, and Laura Schäfer. 2021. *Global Climate Risk Index 2021: Who Suffers Most from Extreme Weather Events? Weather-Related Loss Events in 2019 and 2000 to 2019*. Bonn, DEU: Germanwatch.

Figueroa-Lazu, Damayra I., Jennifer Hinojosa, and Yarimar Bonilla. 2022. *Puerto Rico's 2020 Race/Ethnicity Decennial Analysis*. New York: Center for Puerto Rican Studies at Hunter College.

Hillberry, Russell, and Manuel I. Jimenez. 2024. "The Effect of the Jones Act on Puerto Rico." Research Briefs in Economic Policy No. 380. Washington, DC: Cato Institute

Jensen, Eric B., Anthony Knapp, Heather King, David Armstrong, Sandra L. Johnson, Larry Sink, and Esther Miller. 2020. "Methodology for the 2020 Demographic Analysis Estimates." Washington, DC: US Census Bureau, Population Division.

Langkjaer-Bain, Robert. 2018. "Data, Debt and Disaster in Puerto Rico." *Significance* 15 (1): 10–11. https://doi.org/10.1111/j.1740-9713.2018.01103.x.

Parks, Robbie M., Vasilis Kontis, G. Brooke Anderson, Jane W. Baldwin, Goodarz Danaei, Ralf Toumi, Francesca Dominici, Majid Ezzati, and Marianthi-Anna Kioumourtzoglou. 2023. "Short-Term Excess Mortality Following Tropical Cyclones in the United States." *Science Advances* 9 (33): eadg6633. https://www.doi.org/10.1126/sciadv.adg6633.

Perreira, Krista, Rebecca Peters, Nicole Cafarella Lallemand, and Stephen Zuckerman. 2017. *Puerto Rico Health Care Infrastructure Assessment: Site Visit Report*. Washington, DC: Urban Institute.

Schwabish, Jonathan. 2021. *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*. New York: Columbia University Press.

Stolyar, Lina, Kendal Orgera, and Robin Rudowitz. 2021. "Challenges in the US Territories: COVID-19 and the Medicaid Financing Cliff." San Francisco: KFF.

US Census Bureau. 2020. *Understanding and Using Puerto Rico Community Survey Data: What All Data Users Need to Know*. Washington, DC: US Census Bureau.

———. 2022. "Methodology for the United States Population Estimates: Vintage 2022." Washington, DC: US Census Bureau.

———. n.d. "American Community Survey: A Comparison of the American Community Survey and the Puerto Rico Community Survey." Washington, DC: US Census Bureau.

US Department of Justice. 2007. *Report by the President's Task Force on Puerto Rico's Status*. Washington, DC: US Department of Justice.

Vargas-Ramos, Carlos, Laura Colón-Meléndez, Jorge Soldevilla-Irizarry, Damayra Figueroa-Lazu, Jennifer Hinojosa, and Yarimar Bonilla. 2023. *Pervasive Poverty in Puerto Rico: A Closer Look*. New York: Center for Puerto Rican Studies at Hunter College.

Chapter Four

The Complexity of Data Journalism Education in Indonesian Universities

CHAPTER FOUR

The Complexity of Data Journalism Education in Indonesian Universities

UTAMI DIAH KUSUMAWATI

YEARRY PANJI SETIANTO

Media companies around the world and in Southeast Asian countries like Indonesia have recognized the persuasive power of statistics and data in news stories. But while data collected and published by governments, research institutions, international organizations, and other entities can represent truth and facts, they can also be used to promote lies and misinformation. To counteract this harm—and better understand and communicate people's lived experience—journalists must know how to collect and analyze data accurately and present fact-based stories and narratives to their readers. In this sense, learning data journalism has become essential for Indonesian reporters.

Scholars have different understandings of what data journalism is, but most agree it is a new form of journalism that requires skills in "acquiring, cleaning, analyzing, and presenting data" (Lewis, McAdams, and Stalph 2020, 16). By this definition, data journalism is about developing and using analytical and critical skills to understand data; this differs from computational journalism, which relies heavily on the application of automated journalism and machine learning. Moreover, Brant Houston argues that data journalists do more than browse the internet and write stories; they access databases and apply critical thinking to "provide context and depth to [their] daily stories" (Houston 2004, 3).

Data journalism offers an opportunity for reporters to look beyond available data, particularly data provided by local, provincial, and federal governments. Data journalism education teaches budding journalists to not only examine existing datasets but find alternative datasets to verify claims and establish context.

Positionality Statements

Utami Diah Kusumawati is a researcher and lecturer deeply embedded in the field of data journalism. Growing up in Indonesia and working as a journalist for media like CNN Indonesia and *The Jakarta Post*, she witnessed the challenges and disparities faced by minorities. She also observed a gap between marginalized voices and the data that represent them. Her academic journey in the US and her Indonesian heritage led her to navigate and critically observe the dynamics of data representation and data access. However, she recognizes that not all the things she learned in the US could be applied to the Indonesian experience. The two countries have large gaps in the implementation and the human resource skills of technology in media as well as the level of data literacy among their populations.

Yearry Panji Setianto is an Asian man with a background in US-based university education. He holds a privileged position as a lecturer and researcher specializing in media and journalism. Setianto's participation in Journocoders Indonesia has ignited a strong interest in data journalism, expanding his impact in both academic and non-academic circles on this topic.

The goal is to train reporters to critique data and develop an instinct for identifying misinformation. Rather than simply accepting the figures published by governments, good data journalists will pursue raw datasets directly from data sources and data producers.

Carefully interrogating government datasets is a big challenge for journalists in Indonesia because many government datasets are not easy to find or in machine-friendly formats; that is, they are available only as printed documents or as online images from which it is not easy to extract the actual data. This challenge forces many reporters to abandon stories because of a lack of accessible information or the time-consuming process of manually extracting the data. There is also the problem of overlapping datasets and conflicting data between the central government and local governments—an issue that became clear during the COVID-19 pandemic when, in July 2020, the number of people who died of COVID-19 in Indonesia was estimated to be almost five times higher than the official government count.[1]

The growing importance of data journalism to find and use high-quality data in Indonesian news highlights the need for higher education institutions to adapt how they teach journalism skills in the 21st century. But universities in Indonesia are slow to meet this need. In this essay, we discuss the factors that contribute to this hesitation and the challenges Indonesian universities face to broaden their course and degree offerings.

The Rise of Data Journalism in Indonesia

Indonesian data journalism courses and programs have been growing since the early 2010s, but their adoption has been relatively slow. During this period, only a few media companies experimented with using data in their standard reporting practices because journalists were initially unenthusiastic about this new approach. According to Adi Marsiela, data journalism trainer at the Alliance of Independent Journalists, the aspiration to integrate data journalism into the newsroom came from young reporters who had learned how to work with data at universities in foreign countries, such as the United States and the United Kingdom. Other young journalists were reluctant to write data-driven stories because they were not familiar with computer software and statistics, and more senior journalists trusted the traditional methods of interviewing sources and collecting government data.[2]

On a broader scale, some editors-in-chief were slow to adopt data journalism because several legacy Indonesian media companies were already struggling with transitioning from primarily print platforms to primarily digital platforms. The communications, information, media, and entertainment industries were occupied with accommodating this move toward digital platforms, including developing the platforms, providing technical solutions, and creating wide-ranging online revenue models to attract advertisers (Karimi and Walter 2015). As such, they did not pay much attention to data journalism practices.[3]

In *Diffusion of Innovations*, Everett Rogers explores how rapidly innovation is adopted and how it is influenced by several factors, including compatibility—how an innovation is perceived as being compatible with the "existing values, past experiences, and needs of potential adopters" (Rogers 1983, 15). Data journalism brought transformational change to the traditional ways of reporting, moving from talking to people to analyzing data (Thomond and Lettice 2002). But many journalists debated whether this new approach was compatible with the traditional newsroom culture and whether it would further transform journalism practices.[4]

Despite the generational tensions in the newsroom and the need for reporters to learn new skills, various media companies have tried to incorporate data journalism into their practices in recent years. Reporters are also starting to realize the impact of data-driven stories on the public, mainly in improving the quality of the information and countering disinformation. For these reasons—and because most media companies do not provide official data journalism training—many reporters join workshops provided by such organizations as Journocoders Indonesia, Indonesian Data Journalism Network, and the Alliance of Independent Journalists (see, for example, Cahyarani and Iskandar 2021). According to Reza Khoerul Iman, data and social media journalist at BandungBergerak.id,[5] some local journalists are enrolling in online courses as well (see also Rachmatsyah and Baksin 2023).

As data journalism expands across Indonesian media companies, journalism professors are looking for ways to align their teaching materials and approaches to this new form of journalism. As of 2023, however, Indonesian universities still lack the variety of courses and degrees offered in universities in other parts of the world, especially those in the US and Europe.

Early Adopters and the Issue of Data Journalism Competencies

Universitas Multimedia Nusantara (UMN) and Universitas Indonesia (UI) were early adopters of data journalism curriculum. Their experiences demonstrate the challenges schools face as they try to adjust to the changing media landscape.

In 2017, the digital journalism department at UMN (formerly the journalism department) began teaching a data journalism course to catch up with the global media industry. Lecturers introduced new teaching materials on such topics as data finding, data cleaning, data analysis, data visualization, interactive storytelling, and digital fact-checking. At the time, the department had only two instructors to teach the course: a full-time lecturer and a media practitioner.

Adi Wibowo Octavianto, then the deputy head of the digital journalism department, decided to develop a data journalism course because it aligned with the department's mission to teach a new form of journalism. He researched several data journalism curricula and deemed Northwestern University's curriculum to be the most comprehensive.[6] His lesson plan was a mixture of the Northwestern syllabus and the Microsoft Excel steps for a data storytelling process as suggested by the International Consortium of Investigative Journalism, which included cleaning, sorting, filtering, and analyzing data. For data visualization, he chose the browser-based tool Flourish, which can produce a range of graphs, charts, and animated data visualizations.

Octavianto is one of the first full-time lecturers to teach data journalism in Indonesia, and he said he learned data journalism, Microsoft Excel, and computer programming all on his own. Throughout the semester of teaching the course, he refined his lesson plan to include materials from the *Data Journalism Handbook*, edited by Bounegru, Chambers, and Gray (2012), and other sources. By the end of 2017, the department had officially named the course "Interactive Data Journalism." In 2018, UMN hired two more full-time lecturers. Today, UMN's digital journalism department has three full-time lecturers and invites data journalism experts from around the world to share their experiences. A new data journalism course, "Data-Driven Storytelling," is now required at UMN and is taught to third-semester students.

While many data journalism courses in universities around the world instruct students on using more advanced computer programming languages, like Python or R, UMN decided not to teach those languages, instead focusing on helping students develop the skills to navigate datasets using basic tools, such as Microsoft Excel and Google Sheets, along with other supplementary software, such as Flourish and QGIS. These basic data journalism skills and tools are more accessible and better aligned with the needs of data journalists in Indonesia.

At the state-run UI, data journalism has been offered as an elective course since 2015. Whisnu Triwibowo, head of the department of communication, said the expected learning outcomes for the course include basic knowledge of data journalism techniques, understanding of the practice of data journalism, and familiarity with analytic tools like R.[7] Because of the limited number of faculty members with data journalism background, UI relies completely on practitioners to teach the course. Previously, UI had invited renowned data journalists as instructors, including Wahyu Dhyatmika from Tempo Media Group and Aghnia Adzkia from BBC East Asia.[8]

While the number of university data journalism programs and classes is growing throughout Indonesia, finding practitioners to teach the courses in a university setting can be difficult. Pandan Yudhrapramesti, chair of the journalism department at Universitas Padjadjaran, explained that not many journalists are prepared and able to teach data journalism at the university level, and senior university lecturers struggle with learning the technical side of data journalism, such as coding languages and machine learning.[9] Moreover, most lecturers are busy with administrative tasks, such as campus accreditation and committee assignments; they often rely on junior lecturers to develop new courses.

To address these issues, some universities like Universitas Padjadjaran have decided to collaborate with the Alliance of Independent Journalists, which provides intensive data journalism trainings for students and lecturers, especially younger lecturers. All who complete the training are expected to share what they learned with their peers at their university. In 2020, Journocoders Indonesia, a community for data journalists and data enthusiasts based in Jakarta, provided trainings on developing a data journalism curriculum to university lecturers. Faculty members from 29 different universities attended those training sessions.

Inadequate Data Literacy among Journalism Students

Another challenge for universities incorporating data journalism curriculum is teaching students to become comfortable with using data. In journalism classes, students learn how to write and craft stories based on information from public sources. In data journalism classes, students learn how to use software programs to analyze data, both numeric and text-based, and turn their analyses into stories. This new approach to reporting is difficult to learn, especially for students who are uncomfortable with math or lack sufficient training in math and statistics necessary to conduct analyses.

Universities have taken different approaches to teaching these analytical skills. Anastasya Andriarti, lecturer at Universitas Bakrie, believes it is important to introduce the fundamentals of data journalism like numeric literacy to students early in the course. She allocates at least the first two sessions to general numeric topics so her students will not feel intimidated by data when they begin learning how to craft their stories later in the course.[10] Rossalyn Asmarantika, lecturer at UMN, said the most challenging part of the course for her students is learning how to clean and analyze data.[11] Students tend to rely on basic spreadsheet tools like Microsoft Excel and Google Sheets, instead of carefully interrogating the data through computer code that can be more efficient, scalable, and shareable. To address these challenges, most lecturers provide students with video tutorials on data analysis to watch outside of class; but this requires lecturers to prepare and record materials for asynchronous learning.

Not all journalism is done on a computer, of course. When lecturers at Universitas Padjadjaran observed that students were not able to produce impactful, in-depth, data-driven stories after completing a data journalism course, they reasoned that, because of the COVID-19 pandemic, students had been unable to gain extensive in-the-field training and lacked the requisite interviewing and reporting skills. Students understood the software tools taught in classrooms, but they did not yet have the chance to learn the traditional—and crucial—skills of news reporting.

Instructors across universities are experimenting with different approaches and strategies to teaching data journalism. Many have tasked students to work in groups to highlight the importance of collaborative learning. At UI, groups often consist of students with different majors, including computer science. This strategy enables each student to leverage their unique strengths within the group dynamic. Some students may excel at using tools and software, while others may have outstanding interviewing skills. This collaborative approach to learning is a promising direction for the future of data journalism education in universities.

Conclusion

Integrating data journalism curriculum at Indonesian universities has been challenging, marked by slow progress and various obstacles. The reluctance of newsrooms and classrooms to fully embrace data-driven storytelling reflects an inherent apprehension toward change and the lack of immediate compatibility of the innovative approaches embedded in data journalism with traditional journalism practices. While the emergence of data journalism during the 2010s initially sparked interest, attempts to popularize it faced resistance owing to a range of factors, such as lack of technical familiarity and reliance on conventional reporting methods. Despite these hurdles, some pioneering universities like UMN and UI have incorporated data journalism into their curricula. Their progress underscores the importance of aligning academic standards for reporters with media industry needs.

The challenge of developing adequate data literacy among journalism students remains significant, creating a gap in the newsroom between those who favor traditional reporting and those who prefer data-driven storytelling. As the media landscape continues to evolve in Indonesia, collaborative efforts between universities, media companies, and practitioners are essential to driving effective and transformative change in journalism.

NOTES

1. “Playing with Data,” Editorial, *Tempo*, July 9, 2020, https://en.tempo.co/read/1363234/playing-with-data.
2. Adi Marsiela (data journalism trainer, Alliance of Independent Journalists—the first independent journalists association in Indonesia founded in 1994), interview with the authors, May 2022.
3. Sutta Dharmasaputra (editor in chief, KOMPAS), interview with the authors, October 2022.
4. Afifudin (founder, Digdata.id), interview with authors, March 2023.
5. Reza Khoerul Iman (data and social media journalist, BandungBergerak.id—a nonprofit data-driven media based in Bandung, West Java), interview with the authors, August 2023.
6. Adi Wibowo Octavianto (former deputy head, Digital Journalism Department, Universitas Multimedia Nusantara), interview with the authors, June 2023.
7. Whisnu Triwibowo (head, Department of Communication, Universitas Indonesia), interview with the authors, August 2023.
8. Dhyatmika and Adzkia also worked as guest lecturers at UMN.
9. Pandan Yudhrapramesti (chair, Journalism Department, Universitas Padjadjaran), interview with the authors, August 2023.
10. Anastasya Andriarti (lecturer, Universitas Bakrie—a private university in Jakarta—and data journalism trainer, Alliance of Independent Journalists), interview with the authors, August 2023.
11. Rossalyn Asmarantika (lecturer, Universitas Multimedia Nusantara), interview with the authors, June 2023.

REFERENCES

Bounegru, Liliana, Lucy Chambers, and Jonathan Gray, eds. 2012. *The Data Journalism Handbook 1*. Maastricht, NLD: European Journalism Centre.

Cahyarani, Vania Diah, and Doddy Iskandar. 2021. “Penerapan Citizen Journalism Dalam Pemberitaan Lingkungan Hidup di Media Online.” *Jurnal Riset Jurnalistik dan Media Digital* 1 (2): 71–78.

Houston, Brant. 2004. *Computer-Assisted Reporting: A Practical Guide*, 3rd ed. New York: Bedford/St. Martin’s.

Karimi, Jahangir, and Zhiping Walter. 2015. “The Role of Dynamic Capabilities in Responding to Digital Disruption: A Factor-Based Study of the Newspaper Industry.” *Journal of Management Information Systems* 32 (1): 39–81. https://doi.org/10.1080/07421222.2015.1029380.

Lewis, Norman P., Mindy McAdams, and Florian Stalph. 2020. “Data Journalism.” *Journalism & Mass Communication Educator* 75 (1): 16–21. https://doi.org/10.1177/1077695820904971.

Rachmatsyah, Dimas, and Askurifai Baksin. 2023. “Jurnalisme Data Dalam Media Online.” *Bandung Conference Series: Journalism* 3 (2): 130–35. https://doi.org/10.29313/bcsj.v3i2.8329.

Rogers, Everett M. 1983. *Diffusion of Innovations*, 3rd ed. New York: Free Press.

Thomond, Peter, and Fiona Lettice. 2002. “Disruptive Innovation Explored.” Paper presented at the 9th IPSE International Conference on Concurrent Engineering: Research and Applications, Cranfield, UK, July 27–31.

Biographies and Acknowledgments

EDITORS

Jonathan Schwabish is a senior fellow in the Income and Benefits Policy Center at the Urban Institute. He is also a member of Urban's communications team, where he specializes in data visualization and presentation design. His economic research focuses on nutrition policy and policies that affect people with disabilities. Schwabish is considered a leader in the data visualization field and is a leading voice for clarity and accessibility in research. He has published four books on data communication best practices and is the principal investigator for the *Do No Harm* project.

Wesley Jenkins is a writer and editor on the Urban Institute's communications team. He partners with research teams across the organization to translate their data and analyses into clear, actionable takeaways and ensure that their work is best tailored for use by policymakers, government agencies, and local organizations. Through his work at Urban, Jenkins helps ensure that data visualizations are accessible and accurate, highlighting where data practices may not provide a fully inclusive snapshot of the communities presented.

Shubhangi Kumari is a research analyst in the Metropolitan Housing and Communities Policy Center at the Urban Institute. Her research focuses on participatory methods, community development, housing, and education. Her methodological skills include qualitative and quantitative research, psychometric instrument design, survey research, and program evaluation. Before joining Urban, she worked on gender and sexuality research and on surveys for measuring and improving educational outcomes. Kumari has a bachelor's degree in psychology and applied statistics from the University of Michigan.

AUTHORS

Andrea Berardo leads the research, assessment, and monitoring team at the United Nations World Food Programme for Asia and the Pacific. Previously, he served as a food security and vulnerability analysis expert at the World Food Programme headquarters and its regional bureau for Eastern Africa, and as a technical expert of operation management in the Democratic People's Republic of Korea, Sudan, Tajikistan, and Sri Lanka. Berardo received an MA in political science (with a focus on social research) from the University of Turin, Italy; and an MA in planning for development countries (with a focus on food security analysis) from the Faculty of Planning, IUAV, Italy.

Julia Biedry Gonzalez is a Puerto Rican American technologist and data practitioner who focuses her advocacy work on using data visualization to advance equity. Previously, she served as data consultant and field engineer across leading technology companies and worked with the United Nations, the US government, and leading US and international nongovernmental organizations to help them better understand their data and surface insights crucial to policymaking and advocacy. She is passionate about employing her skills to reduce inequity and has recently published original research in JAMA Network Open about the association of structural fires and heating complaints in New York City, which suggests inequities in safe heating for immigrant communities. Biedry Gonzalez received a BA from Harvard University.

Utami Diah Kusumawati is a lecturer, researcher, and speaker in the journalism department at the Universitas Multimedia Nusantara. She has spoken about data journalism at the Indonesian Statistics Agency, Indonesian Ministry of Communications and Technology Information, Airlangga University, Bina Nusantara University Malang, and Telkom University. Her publications revolve around media adoption of innovation, data for minorities, and public perception toward media reporting. She participated in a research dive sponsored by the UN Global

Pulse Asia Pacific Hub. She received grants two years in a row from the US Embassy to provide workshops and conferences on data and computational journalism in Indonesia and the Ottaway Fellowship to attend the Data Journalism for Investigative Reporting Bootcamp held by the Investigative Reporters and Editors at Missouri University. Kusumawati received an MA in data visualization from the University of Nebraska–Lincoln through a Fulbright scholarship. She also pursued professional education in computational journalism at Columbia University Graduate School of Journalism.

Murali Padmanabhan is a regional disability inclusion advisor at the United Nations World Food Programme for Asia and the Pacific. He is a development professional with 28 years of experience working with disabled people and other marginalized communities, from grassroots to the international level. He promotes empowerment, equality, and mainstreaming through inclusive and participatory program development, policy influence, and advocacy. He is trained in gender, disability, and inclusive development; human rights law; social justice and development law; inclusive project cycle management; and the United Nations Convention on the Rights of Persons with Disabilities. Padmanabhan received an MA in social work from Bharathidasan University.

Anthea Piong is one of a handful of Singaporeans working for the United Nations World Food Programme. After spending some years in Washington, DC, obtaining a master's degree in public policy and working as a data analyst, she entered the field of international development through a Princeton in Africa fellowship in Sierra Leone and has not looked back since. In Sierra Leone, she gained valuable hands-on experience as a monitoring and evaluation officer, collecting and using data to benefit the lives of rural farmers. That experience led her to a role in monitoring with the World Food Programme (WFP) in Bangkok, where she continued using data to measure the organization's impact on food security for vulnerable households in the Asia and Pacific region. Currently, Piong works for WFP in the Afghanistan country office, bringing together information, data, and design to communicate evidence-based humanitarian stories about the work of WFP, inform programmatic decisions, and advocate on behalf of the vulnerable people WFP serves.

Yearry Panji Setianto is an assistant professor in the journalism department at the Universitas Multimedia Nusantara. He teaches classes on communication theories and digital research methods. Previously, he worked in the communication science department at the Universitas Indonesia and the Universitas Sultan Ageng Tirtayasa. His research focuses on the everyday use of digital and social media in Southeast Asia, and he is currently researching the impact of datafication on Indonesian society and exploring how the Indonesian media contributes to the public discourse on artificial intelligence. He has published in the *Journal of Southeast Asian Studies* and *Journal of Media and Religion* and has presented his findings at seminars sponsored by the Association for Education in Journalism and Mass Communication and the Association of Internet Researchers. In 2019, he served as a visiting researcher at the Digital Ethnography Research Centre at RMIT University. Setianto received a PhD in mass communications from Ohio University through a Fulbright Scholarship.

Malvika Sharan is a senior researcher in the Tools, Practices, and Systems program at The Alan Turing Institute. She leads a team of research community managers and co-leads The Turing Way project, which aims to make data science reproducible, collaborative, ethical, and inclusive for researchers around the world. She is a cofounder of OLS (previously Open Life Science), a mentoring and training program that empowers researchers to learn about open science principles, build collaborations with experts, and adopt best practices in their communities. She is an active contributor to several open science initiatives and a Software Sustainability Institute fellow. She has just been named one of the 2024 100 Brilliant Women in AI Ethics™, which recognizes rising stars and pioneers in the AI space. Sharan received a PhD in bioinformatics from the University of Würzburg.

ACKNOWLEDGMENTS

The editors thank Hamutal Bernstein, Jorge Gonzalez, Amanda Hermans, Susan J. Popkin, and Sonia Torres Rodriguez for their thoughtful comments and feedback.

ABOUT THE PROJECT

This report was funded by the Tableau Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization that provides data and evidence to help advance upward mobility and equity. We are a trusted source for changemakers who seek to strengthen decisionmaking, create inclusive economic growth, and improve the well-being of families and communities. For more than 50 years, Urban has delivered facts that inspire solutions—and this remains our charge today. Cover art by Shutterstock.com.

URBAN
INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW | Washington, DC 20024 | www.urban.org