Research Data Governance for Global Health

A Landscape Analysis

Sabrina McCutchan
Duke Global Health Institute
Version 1.1
September 26, 2022

Introduction

The term **data colonialism** has emerged in recent years to describe how private tech companies extract data at large scale from populations in low and middle income countries (LMICs)[†] and use it to drive profits.[1,2,3,4] Data colonialism's two most frequently discussed characteristics are: one, the peoples, communities, and nations who provided data have little to no say in how it is collected or used, and two, they rarely reap its benefits. "Data is the new oil" is a common refrain.[5] As companies and organizations in the tech sector and beyond increasingly rely on data to drive their most essential functions, the parallels to colonialist resource extraction and exploitation become starker.

Global health is also reckoning with colonialist legacies as movements to decolonize global health gain traction.[6] Within academic institutions and non-governmental organizations (NGOs), there is increasing scrutiny of how colonial histories shape institutional processes, curricula, and funding.[7,8,9] Like data colonialism conversations in tech, there is a recurring refrain: under the current system, both decision-making power and benefits are concentrated within high-income country (HIC) institutions.

These two threads converge in questions about how to govern global health research data. Who gets to make decisions about how data is collected and used? Who benefits from data? To adapt an apt analogy from colleagues in South Africa, "if data is the new oil, [data governance] is the new climate change." They go on to define data governance as "the different organising, decision-making, and accountability processes utilised by organisations, companies, local and national governments, and global entities to manage, control, share and exercise power over data."[5] Data governance is closely intertwined with data management, information security, regulatory and legal compliance, analysis and interpretation, and

---

[†] While this report uses the terms HIC and LMIC throughout, we acknowledge there are shortcomings and harms inherent in many of the artificial dichotomies global health uses to describe countries.[65]

funding streams, though these other domains often receive far more attention.

This landscape analysis reviews existing resources about global health research data governance, as well as resources from other fields that might inform approaches to this topic. It is organized into sections that consider data governance rights and impacts by stakeholder group. The appendix briefly summarizes principles and frameworks that have been applied to the topic.

This analysis was prepared as part of the Duke Global Health Institute's (DGHI) Data Governance Initiative. It serves two purposes: to increase awareness of the structural dynamics surrounding global health data governance and how these impact day-to-day research, and to introduce resources for shifting governance practices in alignment with efforts towards decolonization.

Definitions

- **Data colonialism**[4] : When one entity (company, non-profit, donor, government, social enterprise) claims ownership of data that is produced by others or for others, and also takes most of the value of that data.

    - An alternate definition of data colonialism focuses on the rise of big data and individual privacy rights. This definition is used in literature on big data, AI, and surveillance.[10]

- **Data equity**[11] : the consideration, through an equity lens, of the ways in which data is collected, analyzed, interpreted, and distributed.

- **Data governance**[5] : the different organising, decision-making, and accountability processes utilised by organisations, companies, local and national governments, and global entities to manage, control, share and exercise power over data.

- **Data sovereignty**[12] : the control of data flows via national jurisdiction.

- **Data subjects**[13]: persons to who data refer

Data Governance: Rights and Impacts

At a Skoll Foundation conference in 2021, Nexleaf Analytics co-founder and CEO Nithya Ramanathan gave a talk on decolonizing data in which she defined data colonialism and described how to spot it in practice.[4] Her points can be simplified into the categories of rights and impacts:

1. **Rights**: data ownership, access, and uses; how decisions are made *about* data
2. **Impacts**: to whom do positive benefits (financial, intellectual, and operational) accrue, and who might experience negative impacts; how decisions are made *with* data

This analysis considers the rights and impacts at play in global health research data governance, organized into subsections by stakeholder. Stakeholders discussed below include researchers, data subjects, funders, nations, and communities. A given research project will almost certainly involve stakeholders beyond these (see *Figure A*). This landscape analysis focuses on those stakeholders most often discussed.

Defining data.
It has been noted that conversations between stakeholders about data governance are often frustrated by a definitional deficit. One 2022 report found that "references to 'data'… are usually generic and lack a consistent definition or even a shared terminol-

Figure A. Stakeholders in the African data sharing ecosystem. (From Abebe, R., et al. 2021.)

ogy. ...[D]ata could refer to quantitative or qualitative, numbers or narratives, personal or nonpersonal, sensitive or non-sensitive, group or individual, financial, audit or compliance, organisational human resources or beneficiary, situational or contextual reporting, programme-related, as well as indicator data, either aggregated or disaggregated."[14] Data may exist in formats beyond electronic tabular files, such as paper records, biological sam-

Further "Reading"

- "Decolonizing data" with Nithya Ramanathan at the 2021 Skoll World Forum [video recording]. This talk is a succinct, approachable introduction to the realities of global health data governance.

- Narratives and counternarratives on data sharing in Africa, by Abebe et al (2021). This paper uses fictitious vignettes to illuminate the complexities African researchers face navigating the data sharing landscape.

- A Toolkit for Centering Racial Equity Throughout Data Integration, by Hawn Nelson et al (2019). This resource emphasizes equity throughout the data lifecycle. Though written for US-based research, its approaches generalizes to global contexts where other structural imbalances exist.

- Indigenous Data Sovereignty: Toward an agenda, edited by Kukutai and Taylor (2016). This book is considered a seminal work on indigenous data rights, a movement that heavily informs discussions of global data rights more broadly.

- Responsible Data Governance for M&E in Africa. Part 1: Overview of Data Governance, by CLEAR-AA and MERL Tech (2021). This white paper approaches data governance through a legal lens, paying special consideration to data privacy. It also surveys existing legal and regulatory mechanisms in the African context.

ples, genetic data, audio or video recordings, transcripts, medical or electronic device readings, and so on. Clear and shared definitions are essential in conversations about data governance because data ownership, access, and uses may vary greatly depending on the data in question. For example,

there are different factors impacting access to a dataset containing identifiers and a de-identified or anonymized version of the same dataset.

Researchers

Perhaps the most widely understood model of researcher rights is that the researcher generating primary data is its initial owner. Researcher rights fall within the domain of intellectual property law, specifically trade secrets and copyrights. Most primary research data can be considered a trade secret. "...According to international standards, national laws treat information as a trade secret if it derives economic value from not being generally known or readily ascertainable, so long as the information has been subject to reasonable measures to keep it secret."[15] The ability to keep primary data a trade secret is integral to academic publication and grant-writing, two core determinants of career success. Copyright law applies in a more nuanced way, and generally when a researcher seeks to assert "rights of control" over how data is distributed.[15]

Complications emerge when the researcher is an employee because "[e]mployers generally own trade secrets that are developed by their employees within the scope of their employment."[15] Whether employer ownership applies to academic research in the same way it applies in private sectors is a topic of some debate.[16] In practice, institutions settle this question most often in intellectual property policies. A rising number of HIC academic institutions assert institutional ownership through research data policies.[17,18,19] The Research Data Policy proposed by Duke in 2021 took this approach, though no final policy has yet been announced.[20] Guidance from the US Department of Health & Human Services supports such institutional assertions of data ownership.[21]

Questions of ownership are further complicated when more than one researcher and institution are involved. They become especially difficult in international research because intellectual property laws vary across borders. Intellectual property rights might be explicitly recognized, implicitly recognized, or not addressed at all within the various contracts, terms and conditions, and regulatory documentation that govern a research project.

Individuals

The principles underpinning modern individual data rights can be loosely summarized as: individuals should have a choice whether to provide their data, that choice should be informed by an understanding of how their data will be used, and the entities receiving their data should not unilaterally change how data is used after the individual agrees to provide it. A report from the Data Futures Partnership in New Zealand outlines eight key questions individual data subjects have and how organizations can address them (see *Figure B*).[22]

Figure B. Transparent data use dial. (From Data Future Partnership, 2017)

Informed Consent.

Individual data subjects (study participants) also have rights related to research data governance. The right to informed consent has been a cornerstone of clinical research ethics since the Belmont Report. Informed consent practices are coming under increasing scrutiny, especially when collecting data from vulnerable or marginalized populations. They are based on a presupposition of agency—a potential data subject can always decide not to provide their data and suffer no deleterious effects. This may not hold true in practice, such as when receipt of services is, or is perceived to be, conditional on provision of personal data.
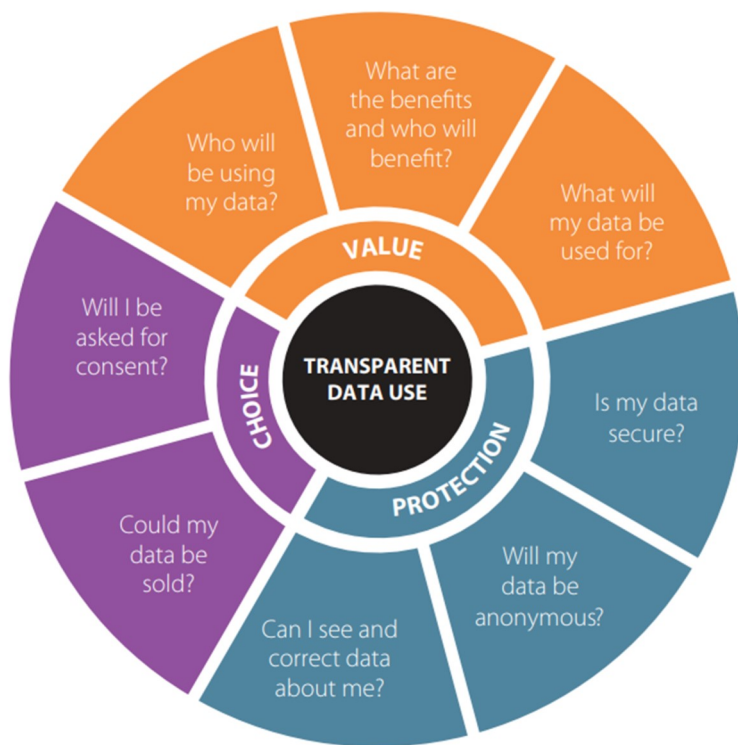
"[B]ecause data has become the cornerstone of public health action and clinical care decision-making, in this evidence-based era the onus is on individuals to disclose information prior to being served. This practice subtly transforms a care process into a transaction contingent upon data acquisition, which at worst introduces coercive and extractive practices into the care necessary to serve clients' vital needs."[23] The more vulnerable the population, the more the data subject's autonomy diminishes. Some researchers have observed that individuals in refugee camps virtually never deny data if they believe providing it might facilitate access to needed services. For this reason data protection offices within major NGOs, including the Red Cross, have signaled a desire to move away from informed consent entirely.[24]

Privacy Law.

In response to the exponential growth of individual data being collected each year, governments around the world are passing new privacy laws. These laws seek to protect an individual right to privacy. The most influential such law in the world today is the European Union's General Data Protection Regulation (GDPR), which came into effect in 2018. The GDPR is a far-reaching security and privacy law governing the personal data of EU citizens and residents. It outlines individual privacy rights, consent processes, data protection principles, security expectations, allowable processing uses, and accountability frameworks.[25,26] A notable provision is the "right to be forgotten," which empowers individuals to request that entities holding their data dispose of it. Many countries have adopted the GDPR as a template for national data protection laws passed since 2018, including Uganda's 2019 Data Protection and Privacy Act and Kenya's 2019 Data Protection Act.[27,28]

Data protection laws passed in HIC counties, like the GDPR, are underpinned by cultural values such as individual autonomy. Some have ques-

tioned how well these adapt to non-Western cultures. Alternative frameworks are being proposed that acknowledge the importance of collective identities in many cultures.[29] Some are based on collective value principles like Umoja (unity), Ujima (collective work and responsibility), and Ujamaa (cooperative economics).[5] Others propose systems for distributing data rights back to community stakeholders, such as through establishing local cooperatives or by licensing certain exclusive rights to local community actors.[2] Such alternative frameworks are only recently emerging.

Funders

The entity financing research may be referred to as a funder, donor, or sponsor. Each connotes different relational terms between the entities providing financial resources and the entities receiving them. Data rights tend to be most clearly defined in commercially sponsored research, where trade secrets and copyright may be shared between the sponsor and researchers (either individual or institutional) or contractually ceded by one of these parties to another.[15] Aside from commercially sponsored research, there is high variation in what type of data rights the financing entity is understood to have and how such rights are exercised. In the following section, all types of financing entities are referred to as "funders" unless otherwise specified.

Formal and Informal Data Requests.

Funders exert control over research data in the sense that they have the option of withholding or discontinuing funding if their expectations about research data are unmet. Sometimes funder data rights are explicitly defined in contracts, such as terms and conditions of award. One contemporary example is the National Institute of Health's new Policy for Data Management and Sharing, coming into effect January 2023.[30] This policy requires the submission of a Data Management and Sharing Plan for all extramural awards and contracts. The plan must address an expectation that researchers "maximize appropriate data sharing." This push towards greater data sharing in HIC scientific research communities is a response to rising concerns about reproducibility, the open science movement, and a reconceptualization of data as public goods.[31,32]

Funders can also make less formal data requests. This frequently occurs in the international aid and development sector as part of monitoring and evaluation. The United Nations, Red Cross, Oxfam, and other NGOs often request aggregate (or summary) research data; requests for individual level data (both identifiable and de-identified) are less frequent. These re-

quests might be to ensure funding is being spent as intended or for "evidence-based decision making" about continuing certain projects. The push towards increasing funder data access is related to corruption concerns and cost-justification for funder oversight bodies.[14]

Funder Power.

Whether they occur formally or informally, it can be difficult to push back on funder data requests. While the policy acknowledges there are valid exceptions and limitations to data sharing, the NIH has signaled that data sharing behavior will increasingly be factored into future award decisions.[33] Several other funders have stated they intend to adopt similar approaches.[34] In one study "[m]ultiple humanitarian interviewees raised the possibility of 'changed' or constrained relationships with donors that could result if they denied a request for data." More established organizations with longer-lasting funder relationships have more agency to negotiate data requests, but these are disproportionately founded and based in HIC countries. Newer local organizations doing the same kinds of work in LMIC contexts are unlikely to have the same advantage.[14]

The amount of control funders exercise over data has drawn scrutiny. In one extreme example, a close relationship between the Gates Foundation and the University of Washington's Institute for Health Metrics and Evaluation (IHME) has been perceived by some as shielding the latter from scrutiny of its data quality and statistical approaches. Some global health actors have expressed reluctance to criticize or challenge the IHME for fear of unfavorable funding outcomes for their own work.[9,35] Even data without quality concerns can be used in ways that negatively impact the people and communities who provided it or whom it is about. In their data rights position statement, Nexleaf Analytics goes so far as to say data should never be used punitively.[36] They assert that decisions to continue, scale up, or stop funding should be decoupled from project data, to avoid incentivizing specific

outcomes over accurate data collection.[4] Funder data requests place additional burdens on researchers which are borne unequally by researchers in higher- and lower- resource contexts. Some requests may be inappropriate.[14]

Funder power is not absolute. Funders rely on implementing researchers and organizations for critical components of the research infrastructure like data collection and statistical analysis. Those funders who have internal expertise don't have it at scale to support their full portfolios. In global health, they also rely on local governments for access to data subjects (though it should be noted jurisdictions are not always recognized in practice). These dependencies create opportunities to push back against funder data requests. Disruptions to existing power dynamics have been proposed. At one panel hosted during the 76th UN General Assembly, panelists proposed that researchers treat national ministries of health (not the funder) as their primary client, and that national governments contribute a higher percentage of research budgets. "If the ministry of health has even… ten percent in that project, then they have some leverage and they have a say in how the money should be spent."[8]

Nations

Data Sovereignty.

"Data is knowledge, is power, is money."[4] Countries increasingly view data generated by, in, or about them as a national resource.[37,38] If data colonialism describes the extractive and exploitative practices of a HIC institution removing data from an LMIC context, some believe its antonym is data sovereignty. Data sovereignty refers to keeping data generated in or from a nation within its domestic legal and regulatory landscape, which implies keeping data on technology resources geographically inside national jurisdiction.[4,12] This is a subject of increasing attention as cloud computing becomes the predominant mode of information storage.[39] If research data is kept in a cloud storage platform for which a HIC institution holds the license, the physical server(s) storing data are almost always located in the same HIC. DukeBox, for example, stores data on the Box corporation's North American servers.

Resources and Capacity.

Justifications for storing data outside the country where it was generated may reference insufficient local resource capacity.[40] HIC based researchers or oversight bodies like IRBs may raise concerns over the resilience and security of LMIC technology solutions. However, greater financial resources do not protect HIC institutions from breaches. Most high-profile cybersecurity attacks on health data have HIC targets, such as the 2017 attack against UK's NHS and the 2022 attack on the Red Cross.[41,42] HIC institutions are more attractive targets for cyber threats due to the volume of valuable data concentrated within them.

Resource differentials are themselves intimately connected to colonial legacies.[6] Others have pointed out that relying on HIC technology resources creates a chilling effect on local innovation in LMICs, which exacerbates existing resource inequalities.[9,43] Focus on the resource gap has itself been criticized as advancing "deficit narratives" that do not serve

LMIC stakeholders. 'In anchoring narratives around these concerns [about insufficient technological resources and data knowledge/training] — which are not unique to Africa and do not capture the full scope of its people's experiences — the strength, agency, and scientific and cultural contributions of communities within the continent are minimized. Community norms, values, and traditions are overlooked, and policies and practice around improving data sharing focus on "fixing" the continent's problems.'[29]

Indigenous Data Governance.

The Indigenous data sovereignty movement gained widespread attention in 2016 with the publication of *Indigenous Data Sovereignty: Toward an agenda*.[44] Indigenous data sovereignty refers to "the right of Indigenous Peoples to own, control, access and possess data that derive from them, and which pertain to their members, knowledge systems, customs or territories."[45] Indigenous data governance, a related concept, is concerned with data governance policies and practices within both Indigenous communities and external institutions stewarding Indigenous data. Emphasis is placed on decision-making power over data belonging to the Indigenous Peoples from which data were derived.[46]

Various advocacy networks for Indigenous data rights exist within HICs. In 2019, the Global Indigenous Data Alliance (GIDA) was founded to unite their efforts and advance Indigenous control of Indigenous Data in the realms of data sovereignty and governance globally.[47] The CARE Principles for Indigenous Data Governance resulted from a 2018 collaborative workshop attended by many GIDA members.[48] Advocacy efforts by the United States Indigenous Data Sovereignty Network (USIDSN) have gained the attention of policymakers. Their work spurred supplementary guidance to the NIH Policy for Data Management and Sharing for American Indian/Alaska Native participant data that recognizes these communities' wishes to "mange, preserve, and share their own data."[49]

Indigenous data rights work has been at the forefront of reconceptualizing legal and ethical bases for data rights at the national level. The definition of Indigenous data does not only include human subjects data about Indigenous Peoples, but also data about their resources and environments and data about nations.[50] As USIDSN notes, data sovereignty in this context is conceptualized differently than in the mainstream. Here data sovereignty and associated rights are derived from the data provider(s): "When data are collected from the people and communities of an Indigenous nation, the data come under the control of that Indigenous nation."[51] The United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) is cited as an international legal basis for work in Indigenous data sovereignty and governance.

Communities

Boundaries delineating a "community" can be drawn in an infinite number of ways. A community may be defined by persons sharing a geographic region, gender, race, ethnicity, tribe, religion, cultural identity, or other characteristics or combinations thereof. For purposes of this landscape analysis, we use "community" in a broad way to discuss collective rights that differ from the national rights (usually exercised by political actors like governments) discussed above.

Data equity in the US.

Within the United States, there is increasing attention to data equity for minority and/or marginalized populations among those conducting community-engaged research. These include interrogations of the ways research is carried out, how data is generated and understood, and how work done with that data impacts communities.[52]

Taking a broad view of the research lifecycle, community stakeholders may have less involvement and influence than other stakeholders even in areas where they are the ones best positioned to understand, interpret, and

Figure C. The inequities wheel. (From Chicago Beyond, 2018).

assess impacts. One report considering researcher, funder, and community stakeholder relationships asks: "What if the structures we use to find what works to improve communities is[sic] negatively impacted by the same
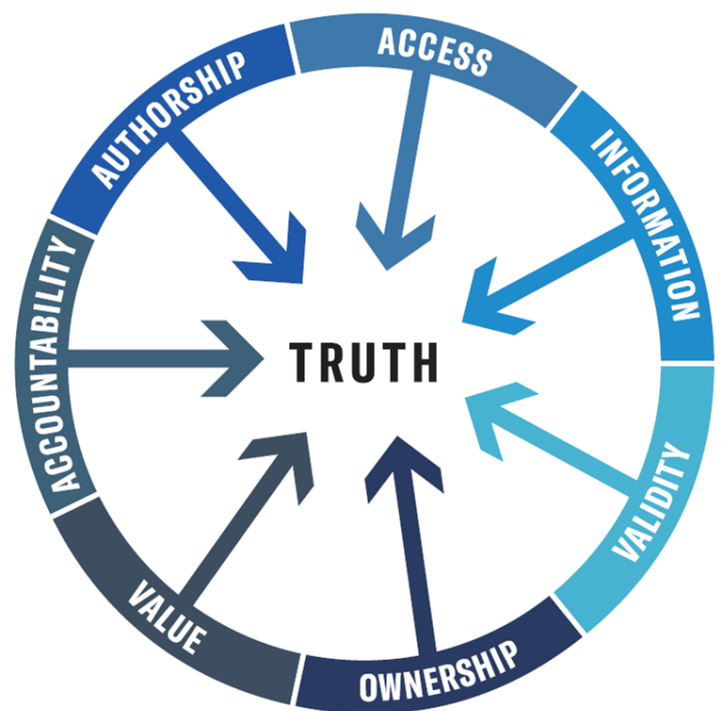
power dynamics that have propped up those systemic injustices?"[53] It breaks this question down into "seven inequities held in place by power": access, information, validity, ownership, value, accountability, and authorship (see *Figure C*). Another report guides researchers to reflect on racial bias and equity in five stages of research: landscape assessment, study design and data collection, data analysis, and dissemination.[54] Like the first, it illuminates how internalized racial biases shape research on numerous levels.

In resources that focus on data systems specifically, the spotlights on community stakeholder input and impacts remain. A report on racial equity in data integrations examines how racial equity can be pursued throughout the data life cycle, including planning, collection, access, processing, analysis, and dissemination. Considerations of benefits and harms arising from data recur as each life cycle stage is discussed. The report also encourages "inclusive participatory governance around access and use," though it does not explore what this might look like.[55] Another report articulates eleven different "drivers" of data systems and how they influence effective community engagement. Examples are organized into six data systems archetypes, ranging from least to most effective at facilitating greater equity.[23] In addition to themes of community exclusion and unequitable impacts, these resources cite biased data as both a cause and product of an unequitable research apparatus.

These resources all highlight how the important context community stakeholders bring to research is often undervalued. Community members are well positioned to understand what questions and programs are most relevant to their communities (the planning stage) and to interpret data about them (analysis, dissemination, and implementation), yet they often have less say than funders or external researchers in shaping research projects and outputs. As a result, they have limited power over how they will be impacted by the research they collaboratively conduct. Communities

that wish to work independently with data face a number of barriers, which might include: poor metadata obfuscating what data are available, opaque avenues for requesting access, costs associated with data access and use, and third-party contractors owning or controlling data.[55]

Data equity beyond the US.

Barriers to working unilaterally with data certainly exist in communities outside the US. One group called Imarisha Naivasha, comprised of stakeholders in environmental resources around Lake Naivasha in Kenya, often participates in generating data about their region but reports routinely being denied access or facing cost-prohibitive barriers to accessing that data.[4] The authors of one fictitious vignette set in Africa imagine how an NGO sharing data generated in a community of minority ethnic group members can have negative effects: first, by publishing results that indirectly identify minority group members, and second, by sharing study data with the national government only instead of also with the community.[29] Such reports suggest a need for greater data sharing back to communities who are providing research data. As one data handbook states, "[w]ithout the data being made 'open' and accessible to [data providers], their information channels about what is happening in their communities might be severely limited... [P]roviding them with proactive access to information is a crucial step towards empowering them to make their opinions known."[56]

However, making data available to communities may not be as simple as it sounds. Working with research data directly requires specific knowledge and skills, and today commonly involves finding, acquiring, processing, and interpreting electronic data files. Competency in tasks such as locating data within an online data repository and processing a data file using specialized software varies hugely across and within communities—local research partners are more likely to have these competencies than a community leader among the study population, for example. Making data

meaningfully available to those outside the research enterprise requires the research team to engage in acts of data translation, transformation, and communication. Mechanisms such as data walks and community advisory boards can bridge this gap.[57,58] These obstacles may be compounded by other community characteristics like low literacy rates, socio-economic disadvantages, and rural/less accessible geographic location.

How Rights and Impacts Intersect

All the above considerations come into play in research data governance. These considerations interact to reveal synergies and deeper complexities, including:

▶ *Shared equity considerations across fields and sectors.*
Equity is framed in terms of ownership and impact when talking about indigenous data rights, community-engaged research in the US, and HIC-LMIC global health research collaborations. Control over, access to, and ability to benefit from data (or at least avoid harm arising from it) disproportionately rests with the stakeholders who occupy a place of structural power.

▶ *Similar losses from community and national exclusions.*
LMIC-based stakeholders (researchers, institutions, governments, and communities), minority community stakeholders, and Indigenous Peoples alike provide valuable context at every stage of the research data life cycle. When their voices and influence are secondary to those of the HIC-based researcher or institution, projects risk asking the wrong questions, making unjustified assumptions, misinterpreting data, and making harmful decisions disproportionately borne by the people and communities being studied.

▶ *Heterogenous rights and impacts within countries.*
Countries are not monoliths. In an international research collaboration, each country represents a plurality of stakeholders whose interests might intersect or diverge to varying degrees. For example, sexual, gender, or ethnic minorities might face persecution from their communities or governments. How should a HIC-based researcher studying such a population navigate discussions about data ownership, ac-

cess, and use with a local community? A collaborating local institution? The national ministry of health?[59]

▶ *Tensions with incentive structures for HIC-based researchers.*
One proposal elevating data sovereignty is to leverage a country's existing data infrastructures for research—for example, by storing data within a national ministry of health's information system.[4] HIC-based researchers might have reservations about this approach, such as getting "scooped" for publication when other scholars access the system, losing access if the system owner terminates collaboration, or technological challenges causing work interruptions. HIC-based researchers may also be wary of being caught between data expectations from HIC actors (their own institution's policies, data sharing expectations from funders and journals) and LMIC actors (LMIC institutional policies, governmental or regulatory bodies) without clear resolution pathways.

▶ *Misalignment of approaches with cultural values.*
Approaches like informed consent and confidentiality safeguards have their roots in culturally Western values like individual autonomy and privacy. When these values do not translate to other cultural contexts one-to-one, approaches predicated on them may encounter some degree of friction. The same may be said of open data and data sharing. One scholar asks: "Should the African continent aspire to develop its own data sharing policies and initiatives grounded in its distinct values, context, and communal culture? This perspective takes data sharing challenges beyond the financial and technological realms and suggests the need for a fundamental rethinking."[29] Others have pointed out manifold complications and difficulties arising from pushing for open data in Africa without considering local research environments.[60]

Appendix: Principles and Frameworks

FAIR Principles.[61]

The FAIR Principles support increased re-use of research data. They articulate what qualities optimize data for re-use. There are four FAIR Principles: Findability, Accessibility, Interoperability, and Reuse. These principles "emphasise machine-actionability" and the encoding of data and associated metadata in machine-readable forms. In scientific research communities, the FAIR Principles are central to ongoing movements towards open data, reuse, and reproducibility.

CARE Principles.[48]

The CARE Principles for Indigenous Data Governance support "Indigenous Peoples' ability to exercise their individual and collective rights to self-determination." They articulate what qualities of human engagement with data support this end. There are four CARE Principles: Collective Benefit, Authority to Control, Responsibility, and

Figure D. CARE Principles logo.

Ethics. These principles are "people and purpose-oriented." The CARE Principles are central to ongoing movements towards restoring data rights to Indigenous Peoples and communities and have been implemented across a variety of research disciplines.
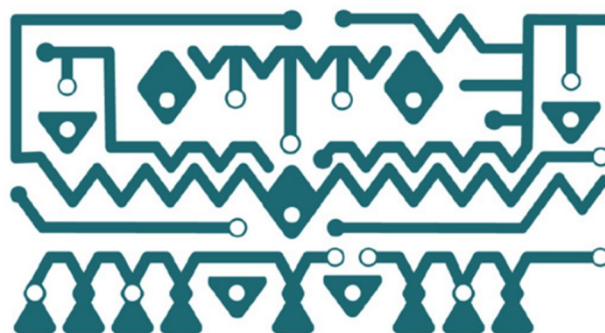
Data Equity Framework.[62]

The Data Equity Framework is a systematic process for examining data projects through an equity lens. It articulates 7 stages of a project: Funding, Motivation, Project Design, Data Collection & Sourcing, Analysis, Interpretation, and Communication & Distribution.

This framework "allow(s) [individuals] to identify and understand each place in [their] work where [they] are embedding a worldview or prioritizing a lived experience." The Data Equity Framework is published by the We All Count Project for Equity in Data Science.
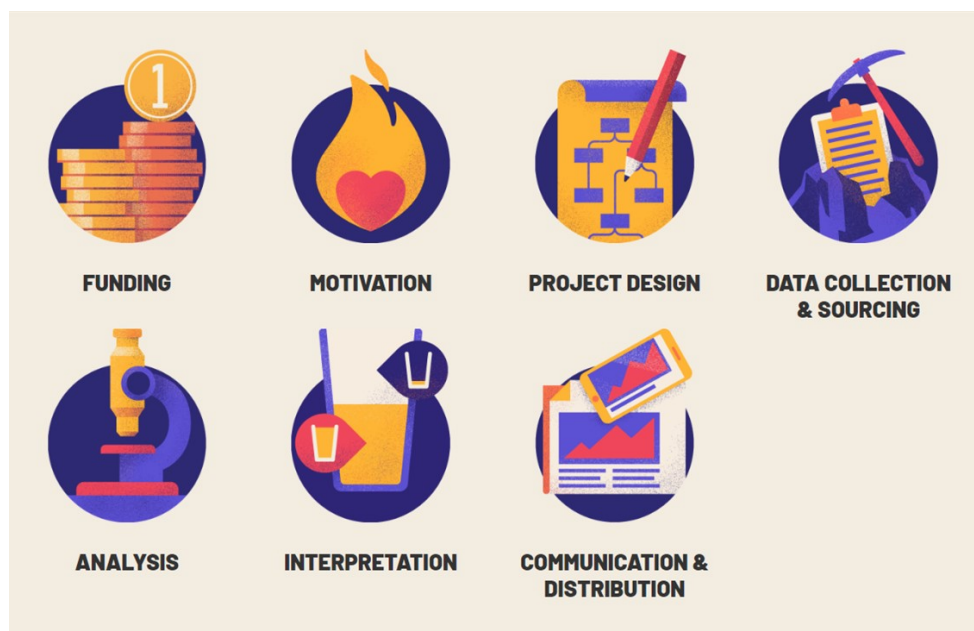

Figure E. The Data Equity Framework (from We All Count).

Data Ethics Canvas.[63]

The Data Ethics Canvas is a tool that "provides a framework to develop ethical guidance that suits any context". It considers data ethics in 3 stages: stewarding data (collecting, maintaining, and sharing it); creating information from data; and decisions about how to act upon data. The canvas articulates questions in 10 domains, including: data provenance, data rights, limitations in data sources, ethical and legislative contexts, purpose, positive and negative effects, and engagement. The Data Ethics Canvas is published by the Open Data Institute.

Principles for Advancing Equitable Data Practice.[64]

The Principles for Advancing Equitable Data Practice is a brief whitepaper that applies the three Belmont Report principles of beneficence, respect for persons, and justice to research data practices. It provides examples of applying these principles in four data lifecycle stages: acquisition, processing and analysis, dissemination, and disposition. Principles for Advancing Equitable Data Practice is published by the Urban Institute.

CIE Data Equity Framework.[23]

The Community Information Exchange (CIE) Data Equity Framework is a tool designed to "guide institutions away from systems that perpetuate harmful practices and towards anti-racist systems that empower communities." It considers the nature of institution-community engagement in areas such as data use, community or individual awareness and involvement, and accountability. It categorizes data systems into 6 design types: predatory engagement, passive engagement, savior-designed engagement, ally-designed engagement, equity-empowered engagement, and liberatory model. These systems are described in terms of 11 "drivers," or elements that comprise the data system. Examples of drivers include system design & governance, informed consent, data stewardship, technology, and security. The CIE Data Equity Framework is published by 211/CIE® San Diego.

References

1 Danielle Coleman, Digital Colonialism: The 21st Century Scramble for Africa through the Extraction and Control of User Data and the Limitations of Data Protection Laws, 24 Mich. J. Race & L. 417 (2019). Available at: https://repository.law.umich.edu/mjrl/vol24/iss2/6

2 Singh, P. J. (2019). Data and Digital Intelligence Commons (Making a Case for their Community Ownership). Data Governance Network Working Paper 02. https://itforchange.net/sites/default/files/1673/Data-commons.pdf

3 Birhane, A. (2020). Algorithmic Colonization of Africa. 17:2 SCRIPTed 389 https://script-ed.org/?p=3888 doi: 10.2966/scrip.170220.389

4 Ramanathan, N., Fruchterman, J. (2021, April 13-15). Decolonizing data [Conference presentation]. 2021 Skoll World Forum. https://skoll.org/session/skoll-world-forum-2021/decolonizing-data/

5 Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA) and MERL Tech (2021) Responsible Data Governance for Monitoring and Evaluation in the African Context. Part 1: Overview of Data Governance and Part 2: Guidance for Responsible Data Governance in Monitoring and Evaluation. Faculty of Commerce, Law and Management | University of the Witwatersrand, Johannesburg, South Africa. https://www.im-portal.org/help-library/responsible-data-governance-for-me-in-africa-part-1-overview-of-data-governance

6 Khan M, Abimbola S, Aloudat T, et al. Decolonising global health in 2021: a roadmap to move from rhetoric to reform. BMJ Global Health 2021;6:e005604.

7 Hirsch, L. A. (2021). Is it possible to decolonise global health institutions? The Lancet (British Edition), 397(10270), 189-190. https://doi.org/10.1016/S0140-6736(20)32763-X

8 Ramanathan, N., Osofisan, W., Eguegu, O. (2021, September 22). What will it take to decolonize aid and development? [Conference presentation]. Devex @ UNGA 76. https://www.youtube.com/watch?v=L06p4wnsd1U

9 Schwab, T. (2020, December 14). Playing Games With Public Health Data. The Nation, 311(12), 24+. https://www.thenation.com/article/society/gates-covid-data-ihme/

10 Thatcher, J., O'Sullivan, D., Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. Environment and Planning D: Society and Space, 34(6), 990–1006. https://doi.org/10.1177/0263775816633195

11 Lee-Ibarra, J. (2020, July 2). Data Equity: What is it, and why does it matter? Hawaii Data Collaborative. https://www.hawaiidata.org/news/2020/7/1/data-equity-what-is-it-and-why-does-it-matter

12 Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. Big Data & Society, 8(1) https://doi.org/10.1177/2053951720982012

13 Data subjects. (2015). In NIST Computer Security Resource Center. Retrieved from https://csrc.nist.gov/glossary/term/data_subjects

14 Fast, L. (2022). Data Sharing Between Humanitarian Organisations and

Donors: Toward Understanding and Articulating Responsible Practice. NCHS Paper 06, April. Bergen: Norwegian Centre for Humanitarian Studies. https://www.im-portal.org/help-library/nhcs-data-sharing-between-humanitarian-organisations-donors-responsible-practice

15 Carroll M. W. (2015). Sharing Research Data and Intellectual Property Law: A Primer. PLoS biology, 13(8), e1002235. https://doi.org/10.1371/journal.pbio.1002235

16 World Intellectual Property Organization. (n.d.) Universities and Intellectual Property. Retrieved June 2, 2022, from https://www.wipo.int/about-ip/en/universities_research/

17 Yale University. (2022). What does it mean that the University has ownership of research data? https://research.yale.edu/faq/what-does-it-mean-university-has-ownership-research-data

18 Columbia University. (n.d.). RCR Data Acquisition and Management. Retrieved June 2, 2022 from https://ccnmtl.columbia.edu/projects/rcr/rcr_data/foundation/index.html#2

19 Rice University. (2012, April 2). Research Data Management. https://policy.rice.edu/308

20 Duke Research & Innovation. (2022). Research Data Initiative. https://research.duke.edu/research-data-policy-initiative

21 Office of Research Integrity. (n.d.). Data ownership. U.S. Department of Health and Human Services. Retrieved June 6, 2022 from https://ori.hhs.gov/

content/Chapter-6-Data-Management-Practices-Data-ownership

22 Data Futures Partnership. (2017, August). A Path to Social Licence: Guidelines for Trusted Data Use. https://aisp.upenn.edu/wp-content/uploads/2019/08/Trusted-Data-Use_2017.pdf

23 Boyd, R., Grounds, K., Johnson, B., Suarez, R., Harrell, T., Rogers, A., Morrison, A. (2021). Leveraging Community Information Exchanges for Equitable and Inclusive Data: The CIE® Data Equity Framework [White paper]. California Children's Trust, 211/CIE® San Diego, Health Leads. https://ciesandiego.org/wp-content/uploads/2021/12/CIE-Data-Equity-Framework-FINAL.pdf

24 Campo, S., Doornbos, A., Gibaudan, F., Hanché, V. (2020, November 2-3). How can you truly manage informed consent in practice? [Conference presentation] GeOnG 2020, Chambéry, France. https://www.youtube.com/watch?v=jIeB-i5fWR0&list=PLDuxmnTc4fTroeNltntD8HaYrs6xHiuBG&index=6

25 Intersoft Consulting. (n.d.) General Data Protection Regulation. Retrieved May 25, 2022 from https://gdpr-info.eu/

26 GDPR.EU. (2022). What is GDPR, the EU's new data protection law? https://gdpr.eu/what-is-gdpr/

27 Uganda Legal Information Institute. (2019, May 3). Data Protection and Privacy Act, 2019. https://ulii.org/akn/ug/act/2019/9/eng@2019-05-03

28 DLA Piper. (2021, December 9). Data Protection Laws of the World: Ken-

ya. https://www.dlapiperdataprotection.com/index.html?t=law&c=KE

29 Abebe, R., Aruleba, K., Birhane, A., Kingsley, S., Obaido, G., Remy, S. L., & Sadagopan, S. (2021, March 1). Narratives and counternarratives on data sharing in africa [Conference paper]. ACM FAccT 2021. https://doi.org/10.1145/3442188.3445897

30 National Institute of Health. (2020, October 29). Final NIH Policy for Data Management and Sharing. Notice number NOT-OD-21-013. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

31 Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). https://doi.org/10.1038/533452a

32 Taylor L. 2016 The ethics of big data as a public good: which public? Whose good? Phil. Trans. R. Soc. A 374: 20160126. http://dx.doi.org/10.1098rsta.2016.0126

33 Jorgenson, L. (2021, April 28). Goals for the NIH Data Management and Sharing Policy [Conference presentation]. Changing the Culture of Data Management and Sharing, United States. https://www.nationalacademies.org/event/04-29-2021/changing-the-culture-of-data-management-and-sharing-a-workshop

34 Franko, M. (2021, April 29). Session V: Encouraging Data Sharing Outside of Mandates – Franko [Conference presentation]. Changing the Culture of Data Management and Sharing, United States. https://www.nationalacademies.org/event/04-29-2021/changing-the-culture-of-data-management-and-sharing-a-workshop

35 Mahajan, M. (2019), The IHME in the Shifting Landscape of Global Health Metrics. Glob Policy, 10: 110-120. https://doi.org/10.1111/1758-5899.12605

36 Nexleaf Analytics. (2022). Data Rights. https://nexleaf.org/data-rights/

37 Press Trust of India. (2019). Data a 'new form of wealth', needs to be taken into account: India. Business Standard. Retrieved from https://www.business-standard.com/article/current-affairs/data-a-new-form-of-wealth-needs-to-be-takeninto-account-india-119062801483_1.html

38 TNN. (2019). 'PM Modi won't compromise on data sovereignty'. The Times of India. Retrieved from https://www. business-standard.com/article/current-affairs/data-a-new-form-of-wealth-needs-to-be-taken-into-accountindia-119062801483_1.html

39 Irion, K. (2012). Government cloud computing and national data sovereignty. Policy and Internet, 4(3-4), 40-71. https://doi.org/10.1002/poi3.10

40 Menon, R. (2017, September 27). Data ecosystems for sustainable development. United Nations Development Programme. https://www.undp.org/publications/data-ecosystems-sustainable-development

41 Bernard, R., Bowsher, G., Sullivan, R. (2020) Cyber security and the unexplored threat to global health: a call for global norms, Global Security: Health, Science and Policy, 5:1, 134-141, DOI: 10.1080/23779497.2020.1865182

42 Mclaughlin, J. (2022, January 20). Cyberattack on Red Cross compromised sensitive data on over 515,000 vulnerable people. NPR. https://www.npr.org/2022/01/20/1074405423/red-cross-cyberattack

43 Mahajan, M. (2018) Philanthropy and the nation-state in global health: The Gates Foundation in India. Global Public Health, 13:10, 1357-1368. https://doi.org/10.1080/17441692.2017.1414286

44 Kukutai, T., Taylor, J. Indigenous data sovereignty: Toward an agenda. Canberra: ANU Press, 2016

45 International Work Group for Indigenous Affairs (IWGIA). (2021, April 28). Indigenous Data Sovereignty. In D. Mamo (Ed.), The Indigenous World 2020 (pp. 654-662) https://iwgia.org/images/yearbook/2020/IWGIA_The_Indigenous_World_2020.pdf

46 Lovett, R., Lee, V., Kukutai, T,. Cormack, D., Carroll-Rainie, S,. & Walker, J. (2019). Good Data Practices for Indigenous Data Sovereignty and Governance. In Good Data, Daly, A., Devitt, SK,. & Mann, M (eds), Institute of Network Cultures, Amsterdam.

47 Global Indigenous Data Alliance. (n.d.). Purpose. Retrieved May 23, 2022 from https://www.gida-global.org/purpose

48 Global Indigenous Data Alliance. (n.d.). CARE Principles of Indigenous Data Governance. Retrieved May 23, 2022 from https://www.gida-global.org/

49 National Institutes of Health (2022, January 25). Request for Public Comments on DRAFT Supplemental Information to the NIH Policy for Data

Management and Sharing: Responsible Management and Sharing of American Indian/ Alaska Native Participant Data. Notice: NOT-OD-22-064. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-064.html

50 Carroll, S.R. (2020, September 24). Operationalizing the CARE Principles for Indigenous Data Governance in RDM (September 24, 2020) [Webinar]. Network of the National Library of Medicine [NNLM]. https://www.youtube.com/watch?v=ICrZgRBtExA

51 United States Indigenous Data Sovereignty Network. (n.d.). About the US Indigenous Data Sovereignty Network. Retrieved May 26, 2022 from https://usindigenousdata.org/about-us

52 The Urban Institute. (n.d.). Equitable Data Practice. Retrieved June 2, 2022 from https://www.urban.org/elevate-data-equity/equitable-data-practice

53 Chicago Beyond. (2018). Why am I always being researched? https://chicagobeyond.org/researchequity/

54 Andrews, K., Parekh, J., Peckoo, S. (2019, August). How to Embed a Racial and Ethnic Equity Perspective in Research: Practical Guidance for the Research Process [Working paper]. Child Trends. https://www.childtrends.org/publications/a-guide-to-incorporating-a-racial-and-ethnic-equity-perspective-throughout-the-research-process

55 Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., Berkowitz, E., et al. (2020). A Toolkit for Centering Racial Equity Throughout Data Integration. Actionable Intelligence for Social Policy, University of Pennsylvania.

https://aisp.upenn.edu/centering-equity/

56 Antin, K., Byrne, R., Geber, T., van Geffen, S., Hoffmann, J., Jayaram, M., Khan, M., Lee, T., Rahman, Z., Simeoni, C., Weinberg, F., Wilson, C., and Rühling, B. (2016). The Hand-Book of the Modern Development Specialist: Being a Complete Illustrated Guide to Responsible Data Usage, Manners & General Deportment. The engine room. https://responsibledata.io/resources/handbook/

57 Murray, B., Falkenburger, E., and Saxena, P. (2015). DataWalks: An Innovative Way to Share Data with Communities. The Urban Institute. https://www.urban.org/sites/default/files/publication/72906/2000510-Data-Walks-An-Innovative-Way-to-Share-Data-with-Communities.pdf

58 Arnos, D., Kroll, E., Jaromin, E., Daly, H., and Falkenburger, E. (2021). Tools and Resources for Project-Based Community Advisory Boards: Community Voice and Power Sharing Guidebook. The Urban Institute. https://www.urban.org/sites/default/files/publication/104938/tools-and-resources-for-project-based-community-advisory-boards_0.pdf

59 Westphal, F., and Meier, C. (2020). Research on the Specific Risks or Constraints Associated with Data Sharing with Donors for Reporting Purposes in Humanitarian Operations. Global Public Policy Institute. https://www.gppi.net/media/GPPi_DonorDataSharingRisks_Report_August2021.pdf

60 Leonelli, S., Rappert, B. and Bezuidenhout, L. (2018). Introduction: Open Data and Africa. Data Science Journal, 17, p.5. DOI: http://doi.org/10.5334/dsj-2018-005

61 GO FAIR Initiative. (n.d.) FAIR Principles. Retrieved May 23, 2022 from https://www.go-fair.org/fair-principles/

62 Datassist Inc., We All Count. (2021). The Data Equity Framework. Retrieved May 23, 2022 from https://weallcount.com/the-data-process/

63 Open Data Institute. (2021, June 28). The Data Ethics Canvas. https://theodi.org/article/the-data-ethics-canvas-2021

64 Gaddy, M., Scott, K. (2020, June). Principles for Advancing Equitable Data Practice [White paper]. Urban Institute. https://www.urban.org/sites/default/files/publication/102346/principles-for-advancing-equitable-data-practice.pdf

65 Khan T, Abimbola S, Kyobutungi C, et al. How we classify countries and people—and why it matters. BMJ Global Health 2022;7:e009704. doi:10.1136/ bmjgh-2022-009704