Viewpoint

Decolonize Data
By Nithya Ramanathan, Jim Fruchterman, Amy Fowler & Gabriele Carotti-Sha
Decolonize Data

The social sector too often extracts and siloes data from the communities it supposedly serves.

BY NITHYA RAMANATHAN, JIM FRUCHTERMAN, AMY FOWLER & GABRIELE CAROTTI-SHA

The social sector aims to empower communities with tools and knowledge to effect change for themselves, because community-driven change is more likely to drive sustained impact than attempts to force change from the outside. This commitment should include data, which is increasingly essential for generating social impact. Today the effective implementation and continuous improvement of social programs all but requires the collection and analysis of data.

But all too often, social sector practitioners, including researchers, extract data from individuals, communities, and countries for their own purposes, and do not even make it available to them, let alone enable them to draw their own conclusions from it. With data flows the power to make informed decisions.

It is therefore counterproductive, and painfully ironic, that we have ignored our fundamental principles when it comes to data. We see donors and leading practitioners making a sincere move to decolonize aid. However, if we are truly committed to decolonizing the practices in aid, then we must also examine the ownership and flow of data.

Decolonizing data would not only help ensure that the benefits of data accrue directly to the rightful data owners but also open up more intentional data sharing driven by the rightful data owners—the communities we claim to empower.

Recognizing Data Colonialism

In their 2019 book The Costs of Connection, Nick Couldry and Ulises Mejias apply the term “data colonialism” to the practice of extracting data in ways that repeat or mimic historic colonialist practices of extracting natural resources and reinforce the colonial paradigm of exerting decision-making authority over native peoples.

Some of the earliest references to data colonialism we have found come from Indigenous communities, who have long been fighting the profitable extraction and exploitation of local knowledge by others. Today, the Urban Indian Health Institute, a Seattle-based health research center that serves American Indian and Native Alaskan communities, is one of the organizations that are working to “decolonize data, for indigenous people, by indigenous people.”

We see data colonialism in the for-profit sector as well as development organizations. Our working definition of data colonialism is the practice of claiming ownership of data that is produced by others or for others, and appropriating most of the value from that data. While doing exploratory research in Africa for a new tech social enterprise, coauthor Jim Fruchterman interviewed a local leader from Kenya’s Lake Naivasha region, who commented that his ecosystem was one of the most studied in Africa’s Rift Valley. Yet, he and his peers had little access to that data, including weather, water-use, and productivity data, or to the resulting analyses. Instead, such research outputs were locked up in the databases of universities, international NGOs, companies, or public agencies, and were usually inaccessible or unaffordable.

The phenomenon has the following essential elements:

- Data (and learnings) are taken out of countries or communities without the consent or knowledge of the rightful data owner—the country, community, or population where data is collected or about whom the data is collected.
- Knowledge is generated from the data outside of the source country or community, without the rightful data owner’s involvement.
- The data extraction establishes an inherent power imbalance—which either preexisted the data collection or arose as a result of the data collection—between the data collectors and the rightful data owners.
- The collectors create systems and ideologies that reinforce or justify their appropriation of data—in the name of intellectual rigor, as a requirement for monitoring and evaluation, or according to a for-profit business model. Regardless, these systems frequently determine where and how resources are allocated.

From our experience, data colonialism has at least three harmful consequences. First, it generates low-quality decisions. Because analysis happens without community involvement, data colonialism leads to conclusions that are often divorced from
the reality on the ground and the rightful data owners’ perspectives and interests.

Second, it *disempowers*. Data colonialism removes decision-making power from the country or communities directly impacted. For example, one East African leader who worked in their country’s health ministry shared their frustration and humiliation when, while sitting at a conference, they saw data for the first time being presented about their country by a researcher who had traveled there to conduct a study. Not only did the researcher take full credit for this work, but the country’s decision makers were robbed of the opportunity to use the data to navigate and advocate for their own path forward.

Third, it *misappropriates resources*. Data analyses are often used to determine how and where money flows. Consequently, data colonialism cuts out the rightful data owners when funders and decision makers allocate capital and other resources. For example, large-scale health programs have historically been determined by incidence or prevalence data, with large sums of money flowing in and out of communities based on data they did not collect or help to properly contextualize. Mechanisms such as results-based financing can use data to justify withholding funding from a region or program, effectively punishing communities based on analyses in which they did not participate.

**Data as a Reusable Resource**

Traditionally, extractive colonial practices have tended to focus on finite natural resources such as oil and diamonds. The limited quantities and demand for these resources increased their value. But data is not like oil. It can be replicated and shared without cost. So why are we acting as though it were scarce?

In the private sector, companies retain the great majority of their data in proprietary silos because being the sole owner of information is increasingly the basis of corporations’ competitive advantage. Big tech is incentivized to own and capture data from others, and the massive proprietary databases are a huge source of wealth that accrues to companies.

Data is not a finite resource in the social impact sector, yet that is how we operate. We believe it is this reflexive tendency, rather than an insidious intent to keep communities down, that underpins data colonialism today. Sharing data would accelerate shared impact by aligning incentives and highlighting solutions that truly work. The sharing of aggregated and anonymized data sets can bring clarity to complex environments based on shared insights to work toward continuous improvement to existing systems.

For example, in 2017 government health leaders gathered in Mozambique to discuss vaccine refrigerator performance data that the country’s Ministry of Health was collecting using real-time sensors. Tanzania’s health ministry was considering purchasing some of the same refrigerators. So, Mozambique shared the fridge performance data with Tanzania. Such a simple act highlights the predicament of many other countries whose data is owned by third parties (often companies) and who therefore cannot help themselves or their neighbors this way.

Of course, not all data should be propagated or openly shared. Human rights witness data, child abuse reports, individual health records, and HIV status are all examples of sensitive information about individuals that should not be shared. We are referring to data that can be responsibly used more widely, such as data on equipment and infrastructure, program effectiveness, drug effectiveness, or land use.

**Four Principles for Change**

Data must no longer be extracted or captured from countries and communities. If we in the social sector want to truly engage in systems change in partnership with communities around the world, we need to extend that partnership into the realm of today’s most durable resource: data. We propose the following four principles regarding the ownership, privacy, sharing, consent, and appropriate use of data as guideposts for decolonizing data and shifting power back to the communities where data is collected.

First, data should be owned by the communities and countries where it is collected. Decisions about data use, analysis, access, and interpretation should be driven by these rightful data owners, who should have the tools and resources to interpret and act on the data. Gavi, the Vaccine Alliance, has been an early leader in this shift, by insisting that countries own their data and reiterating the country ownership principle as it applies to data generated by vaccine refrigerators.

Second, the rightful data owners should determine which data should remain private. This principle goes beyond protecting personally identifiable information about individuals, which is increasingly regulated at the national level. It can include sensitive information that the community wants to protect, such as the locations of natural resource deposits or antiquities.

Third, data should be shared only with meaningful consent, for meaningful value. Data is prized because of the ability to use it for rapid learning. The sharing of data, when done in a way that does not increase risk for any individual or community, underpins our ability to maximize the value of that data for the rightful data owners as well as the social impact sector. In this context, meaningful consent and value is derived from respect for the rightful data owners—honoring their cultural norms, ensuring that their learning and use is prioritized, and placing control squarely in their hands. When implemented in this way, this principle should encourage a true value exchange in data sharing, rather than expropriation.

Fourth, data should not be punitive. The modern world uses data to learn and improve, and this use demands quality data. Data that is used to punish creates perverse incentives to generate false data to avoid punishment, which negates the value of data for learning and impact. Moving away from data colonialism is not an easy transition. We recognize the humility, introspection, and commitment it takes to identify and eliminate data colonialism in all of our work. If we can make this shift, we can strengthen collective action and help ensure that the benefits of data directly and immediately accrue to the people and places from which it is collected. This is the purpose of decolonizing data and a core value of the social impact sector.