Beyond Good Intentions: Data Sharing in and for Africa

February 2020

1 Introduction

Data sharing is a complex issue involving multiple stakeholders and models of governance. These issues are important because technologies – including machine learning – are guided by and depend on data inputs and feedback. Whose data we decide to collect, share, and make inferences about expresses beliefs about which populations that are considered worth learning about. In absence of data, progress is slowed in terms of integrating technology into the economy, such as manufacturing or agriculture. On the other hand, when data is collected without knowledge about the population or context from which the data originated, our assumptions about the data might play a larger role in outcomes, especially if we have design our models in a way that applies learning from one context to the assumptions we make about another. In this way, the mannerisms and mechanisms by which data is shared via data processes and collectors (both technical and human) impacts the distribution of economic benefits and opportunities in societies, especially when learning from de-contextualized data allocates those things. [13].

In economies that rely heavily on (information) services or financialization, the ability to accumulate data is an important. Firms capitalize on the data they collect to allocate services or commodities based on data inferences [13]. Those who are able to monopolize and accumulate vast amounts of data are at an advantage in the inference economy. However, outside data extractors may own and determine the data that is available[6]. When data is scarce, access is typically limited (i.e., the data is collected or collect-able, but privately kept), and affordability therefore becomes problematic (i.e., since data is private, access is costly and only available to those who can buy it). Data policies are a mechanism for balancing accessibilty, affordability, privacy and security.[1]

Globalized standards, norms of practice, and regulations for data are not agreed upon [6]. Hence, the practices of data economy stakeholders are sometimes divergent, ad hoc, contradictory, and/or violate community desires or hopes, let alone values and ethics for data-use. In our work, we study this tension in context of data-use in communities across the African continent. Africa, as a continent, is home to diverse and rich communities; histories; traditions; cultures; languages; and values. But in recent years, the African continent, as a whole, has been considered a "frontier" for data opportunists[5, 9]. This, some argue, alludes to a colonialism for powerful multinationals and data brokers who seek a stake in data economies in Africa, where data markets are expected to have the largest worldwide growth in the near future [1, 6].

This narrative speaks to a dynamic we seek to address; namely, how data practices – the collecting, sharing, and ownership of data rights – determine where power is centered (and for

[1] In future work, we are using computing methods to collect, classify and compare data policies from around the globe to compare how those policies enforce or address data issues reported in the news and social media.

whom)? How do histories of uneven power entrench legacies of disparity, colonialisms, ethnocentracisms, slavery, national and racial superiorities – as well as patriarchal values – into data transactions in and for Africa at this time? Histories (re)generate across centuries, through the dynamics of economies, and determine who has power, and who reaps the most rewards in terms of wealth. For this reason, it is our ethical responsibility to interrogate patterns in how power is decided, and to make sure our idea of which things deserve addressing do not replicate historical oppressions that we seek to eliminate from our futures. Part of this power is what we name in our studies. The 54 distinct African countries offer a wealth of previously untapped and potentially valuable data sources. But the history of power in these countries is mirrored by the power dynamics between data collectors and local stakeholders. This fact challenges us to imagine a present, and future, characterized by data colonialism [1, 6]; where African individuals, communities, and entire heterogeneous geographies of people, have their data accessed, harvested and shared, but do not reap the same rewards from data transactions as the extractors/beneficiaries (Benjamin, 2019).

Research Lens

We approach our research from the lens of "humanism" in human-computer interaction [3]. As it is called, humanistic HCI leverages theories and concepts from the humanities for the purpose of building or designing systems and processes [3]. This type of work may take methodologies from the humanities for the purpose and in the service of design research [3]. Accordingly, we leverage interviews and storytelling to diagnose the issues ("economic inequality", "colonialism", "uneven bargaining power") germane to our research question ("how could we collect and share data respectfully?").

Similar acts along the data creation and use pipeline have been studied among other geographic contexts; e.g., Myanmar and in other contexts, including e.g., deploying data to preserve privacy [4]. Others including [11] explore protest uses of technology, in which users actively resist technological exploitation through deleting or intentionally opting not to use software. Other forms of technological protest included installing ad blocking technology to subvert or disrupt business models of companies. While the authors focus on users in the Global North, they also highlight the need to understand protest technology usage in other regions of the world, and why different communities engage in these practices. In our paper, we respond to this call by collecting and analyzing data about data practices (and resistances) in and for communities across the African continent.

In the tradition of critical design, we offer an alternative narrative ("theory of power") to the normative assumption that promotes the idea that data is (equally for everyone) the "new oil" of economic growth and development [8]. We follow the ethos of work in critical design that seeks to "make consumers more critical about their everyday lives, and in particular, how their lives are mediated by assumptions, values, ideologies, and behavioral norms inscribed in designs" [2]. We rebut the trend reflected by what is named in the literature: that data is exclusively a priority of and domain for stakeholders of the Global North. Our central thesis is that historical and contemporary differences in power held among individuals, communities, and nations affect outcomes and who benefits from transactions in the data economy. We argue that historical patterns of oppression, inequality and injustice are reflected in these power differentials, and hence, in data transactions.

2 Methodology

We rely on storytelling – a method employed by many cultural groups within the continent, including in countries like Ethiopia where the tradition has survived for many centuries – to discuss instances when well-intentioned, foreign-based individuals and organizations collect (or request) data that includes details that could violate community norms and standards. In certain contexts, normative practices cause unintended harms that outweigh any benefits of charitable work. The practices may, however, violate group privacy notions that are more common in the African context [12], potentially leading to stigmatization.

3 Preliminary Stories

In this section, we narrate personas that tell stories about data experiences. Personal and collective narratives are powerful, and invite us to change (whether prospective or actions). Stories, through narratives, call attention to social issues while respecting sensitive dynamics that make people vulnerable. Stories allow us to participate, take in broader communities and engage those communities in conversations about justice. Stories tell us about ourselves and others, though sometimes we only approach narratives to learn about the opposite. Stories are powerful for this reason, and because they are able to convey and conceal facts at the same time. Thus, those who are made vulnerable by speaking their truth will find vehicles for truth-telling in them.

Personas

Data Colonialism and the Global North

DrinkUS West LLC is an NGO based in Europe working on sustainable access to water and health. This NGO hopes to support people in Buranda by improving access to potable water. In order to do so, DrinkUS deems that it needs to collect data about people's water use habits and their lived realities. The NGO's researchers decide to use surveys to collect this data from members of the yetet'ebek'e community, the region where DrinkUS will experiment with water accessibility technology and online monitoring of resources.In the survey, researchers ask individual respondents about their age, gender, income, if they own a computer or mobile device (and how many), their relationship status, number of members in their household, if they have children (and how many), how water and utilities are accessed, and their knowledge of potable water accessibility. DrinkUS researchers plan to collect and analyze this data to determine at what price they may offer potable water access to the community, and which technologies could be used to monitor and manage access. They also want to understand the scale and scope of the problem, i.e., how many people face water accessibility issues. Furthermore, to improve completion rates of the survey, DrinkUS incentivize survey takers by paying each participant the equivalent of $20 U.S. dollars in Kwacha, the local currency, for completing the estimated 15-minute long survey. Because DrinkUS researchers want to contact survey participants about future project-related opportunities, they also collect participant contact information such as name, address, and phone number. From the survey data, DrinkUS researchers compile and infer what community members could afford to pay, which members were without access to potable water, and who was at most risk for water-related health impacts or disease. This work is later published in a research report in partnership with the Buranda Health

Ministry.

While DrinkUS's research was well-intentioned, the NGO failed to comprehend the consequences of sharing their research findings publicly, including their predictions about which geographic locations in the yetet'ebek'e community were at risk for water-related disease. Such findings, even shared in aggregate, can reveal information on the community as a whole. While notions of privacy often focus on the individual [7, 10], there is growing awareness that collective identity is also important within many African communities, and that sharing aggregate information about communities can also be regarded as a privacy violation [12]. Global North agents and organizations, whether governments or multinational firms, exert an out-sized influence in the global data economy. For the many African data economies, this can translate to issues around lacking access to data collected by Global North stakeholders, not reaping the same rewards and benefits as non-locals, and having to trade autonomy in decision-making for the marginal financial incentives offered by Global North agents.

The Journey of African Scholars

Soil and apartheid. A PhD candidate from Bozatta conducting research in Nova Africa wants to get soil samples across a few Nova Africa provinces for her research. In her efforts to do so, she learns that farmers are not enthusiastic about sharing this information, but she is unable to understand why. After months of attempts, she eventually learns from her neighbors that farmers may not divulge this information to researchers due to fear that the Government might want to claim ownership of such lands. Even for local researchers, this issue might not seem to be the case. This issue was exacerbated by the apartheid regime, when land grabs were widespread, and black Nova Africans were mostly affected. Understandably when sharing data has led to the forfeiture of rights, access becomes the price of histories, as in most locations it is.

Gender-based violence. A PhD candidate researching the prevalence of gender-based violence in Nova Africa, which is regarded as one of the highest in the world. In conducting this research, she experiences a few unexpected encounters and shares it with her office mates who are from Nova Africa. She tells them that many participants got emotional when sharing previous (or personal) experiences, and asked that the session be discontinued. One participant even questioned why the interview was so sensitive and decided not to share her health status. The participant noted that she might be stigmatized if she revealed her health status. Her office mate tells her that, here in Nova Africa, it is difficult to convince participants to release information, especially if you are not a local Nova African. There is a perceived notion that foreigners are here to steal everything that belongs to them. Hence, the well-articulated narrative of referring to foreigners as ("kwerekwere"/"amakwerekwere"). Her other office-mate advises that in most rural areas, some participants might seek secluded venues that are secure for sharing experiences. The issue of trust is a serious problem in data collection, ownership, and sharing.

Young researchers travelling across different African countries may encounter challenges when trying to collect data. This challenge is particularly well-pronounced when the researcher is not local to the region or community in which the data is being collected. Differences in languages and cultures, and imbalances in trust and power breed suspicion and doubt about motives impeding data access and sharing. With more than 1500 languages being spoken across the African continent, locals residing in rural areas are more likely to communicate in their dialect with little or no understanding of the English. Furthermore, Africa is a continent that is rich in cultures and norms. The expected belief is that a researcher should exhibit an understanding of the culture and norms of the data

collection site when conducting research. Else, they may be perceived as an "out-cast." However, as a non-indigene, understanding the culture of the study site might be daunting, hence, delaying the research output and productivity of the researcher. Despite the abundant and novel sources of data in Africa, this lack of understanding of cultural norms might become a barrier to data access.

Livestocks and Livelihoods

A young researcher in Wakanda receives a two-year grant from a well-respected, family-run foundation from the Global North. The research project outlined an ambitious goal of ushering in prosperity for the livestock farmers by making the mainstream markets work for them. The researcher outlined his intentions to do so by collecting various types of data. These include genomic data of crops and livestock and enterprise data for the farm. The genomic dataset would help determine if the right crops are being planted and the right animal breeds are being reared to support the production system. It would help provide the farmers vital intelligence on susceptibility of their crops and animals to diseases and performance of their farm against aggregate data. The project was structured as a donor-funded research project with clear "research" and data collection and analysis objectives. Yet, no protocols for data storage and scope of use were defined. Furthermore, scant detail was available on if and how the collected data would be used beyond the time period of the donor-funded project.

As expected, the young researcher and his team diligently collect the genomic data as well as the farm data from a cluster of villages in several counties of Southern Wakanda. The data collection process was more challenging than originally anticipated. Even with a local team well-versed in the norms and practices of the region, farmers seemed skeptical to give up their data. Slowly, as the end of the two-year grant period approached, the data collection exercise also came to a close. Encouraged by this newly collected, valuable dataset, the researcher established a for-profit company using the data collected as part of the research project. In its mission statement, the company articulated the need for precision agriculture in today's data-rich environment as an enabler of better agricultural practices. The source data on which the company was built came from that collected as part of the research project. In addition, due to infrastructure challenges, the data is now housed in a data centre at a University in the Eastern United Provinces, an influential country in the Global North which is also home to the foundation that originally funded this research. The company has set its sights on expanding across Wakanda. Even though the trust established with the farmers in Southern Wakanda has been destroyed by the actions of the researcher, he is confident that other farmers will see the potential of his data engine and believe in the power of its predictive analytics.

4 Discussion and Ongoing Work

Without the economic return to communities, whose data gives rise to such wealth, we risk reproducing development trajectories of previous eras, which have caused many African countries to be labeled by the Global North as "developing" rather than "developed". These eras are characterized by those in the Global North as histories dominated by colonialism and exploitation of the continent - and also by - the manifest destiny of the Global North itself. However, this narrative undervalues or ignores the rich, multitude of histories contained and empowered by peoples living all across the continent. Partly, the perspective that Africa (as a continent of 54 countries of many diverse communities) was/is dominated by outsiders, extraordinarily and only impoverished, and

at times, lacked the capacity to develop itself, is a product of our global research community. Our publications center the work of development on what research from the Global North has assumed, and not on what African communities would have necessarily told us, if we had asked or listened.

Our research encourages learning about communities at the local level. The work of understanding data contexts need not take constraints as the exclusive focus. We should account for rich, local capabilities; both in terms of human capital situated in and learnable at the local level, and in terms of local data infrastructures (rule, access, security and privacy mechanisms). Finally, multinational data work must center on the human needs of data subjects. This does not necessarily dictate that outside standards apply to local contexts, especially when those standards make no sense. Toward this end, we encourage conversations that foster data governance norms that are attuned to (but not limited by) legal and technical constraints, while sensitive to the social and cultural-realities in which data is collected and shared. Ultimately, we argue equitable growth requires data infrastructures that are driven by communities. In addition, data equity in the global data economy necessitates community ownership is normalized, encouraged, and that what is local is primary.

References

[1] Birhane Abeba. The algorithmic colonization of africa: Startups are importing and imposing ai systems founded on individualistic and capitalist drives. *Real Life Magazine*, 2019.

[2] Jeffrey Bardzell and Shaowen Bardzell. What is" critical" about critical design? In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3297–3306, 2013.

[3] Jeffrey Bardzell and Shaowen Bardzell. Humanistic hci. *interactions*, 23(2):20–29, 2016.

[4] Finn Brunton and Helen Nissenbaum. *Obfuscation: A user's guide for privacy and protest.* Mit Press, 2015.

[5] Moustapha Cisse. Look to africa to advance artificial intelligence. *Nature*, 562(7728):461–462, 2018.

[6] 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. *Mich. J. Race & L.*, 24:417, 2018.

[7] Eric Horvitz and Deirdre Mulligan. Data, privacy, and the greater good. *Science*, 349(6245):253–255, 2015.

[8] Patrik Hummel, Matthias Braun, Steffen Augsberg, and Peter Dabrock. Sovereignty and data sharing. *ITU Journal: ICT Discoveries.* Πρόσβαση από: *https://www. itu. int/en/journal/002/Documents/ITU2018-11. pdf Ημερομηνία τελευταίας προσπέλασης*, 25(08):2019, 2018.

[9] Hao Karen. The future of ai research is in africa. *MIT Technology Review*, 2019.

[10] Patty Kostkova, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, Peter Knight, Corinne Marsolier, Rachel A McKendry, et al. Who owns the data? open data for healthcare. *Frontiers in public health*, 4:7, 2016.

[11] Hanlin Li, Nicholas Vincent, Janice Tsai, Jofish Kaye, and Brent Hecht. How do people change their technology use in protest? understanding. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–22, 2019.

[12] Moses Namara, Daricia Wilkinson, Byron M Lowens, Bart P Knijnenburg, Rita Orji, and Remy L Sekou. Cross-cultural perspectives on ehealth privacy in africa. In *Proceedings of the Second African Conference for Human Computer Interaction: Thriving Communities*, pages 1–11, 2018.

[13] Babatunde Okunoye and Gbenga Sesan. Data for good in africa–towards increased access and enabling data policies. *ITU Journal: ICT Discoveries*, 1(2):1–4, 2018.