

Narratives and Counternarratives on Data Sharing in Africa

Rediet Abebe
University of California, Berkeley

Kehinde Aruleba
University of the Witwatersrand

Abeba Birhane
University College Dublin & Lero

Sara Kingsley
Carnegie Mellon University

George Obaido
University of the Witwatersrand

Sekou L. Remy
IBM Research - Africa

Swathi Sadagopan
Deloitte

ABSTRACT

As machine learning and data science applications grow ever more prevalent, there is an increased focus on data sharing and open data initiatives, particularly in the context of the African continent. Many argue that data sharing can support research and policy design to alleviate poverty, inequality, and derivative effects in Africa. Despite the fact that the datasets in question are often extracted from African communities, conversations around the challenges of accessing and sharing African data are too often driven by non-African stakeholders. These perspectives frequently employ a *deficit narratives*, often focusing on lack of education, training, and technological resources in the continent as the leading causes of friction in the data ecosystem.

We argue that these narratives obfuscate and distort the full complexity of the African data sharing landscape. In particular, we use *storytelling* via fictional personas built from a series of interviews with African data experts to complicate dominant narratives and to provide counternarratives. Coupling these personas with research on data practices within the continent, we identify recurring barriers to data sharing as well as inequities in the distribution of data sharing benefits. In particular, we discuss issues arising from power imbalances resulting from the legacies of colonialism, ethno-centrism, and slavery, disinvestment in building trust, lack of acknowledgement of historical and present-day extractive practices, and Western-centric policies that are ill-suited to the African context. After outlining these problems, we discuss avenues for addressing them when sharing data generated in the continent.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; • **Social and professional topics** → **Government technology policy**.

KEYWORDS

Data sharing, data access, open data, storytelling, decolonial theory

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

FAccT '21, March 3–10, 2021, Virtual Event, Canada

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8309-7/21/03...\$15.00

<https://doi.org/10.1145/3442188.3445897>

ACM Reference Format:

Rediet Abebe, Kehinde Aruleba, Abeba Birhane, Sara Kingsley, George Obaido, Sekou L. Remy, and Swathi Sadagopan. 2021. Narratives and Counternarratives on Data Sharing in Africa. In *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021, Virtual Event, Canada*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3442188.3445897>

1 INTRODUCTION

Data sharing is often regarded as a crucial element of research and scientific advancements, a hallmark of transparency, and a key to economic growth [14, 21, 26]. Data sharing expands knowledge and contributes to a better science [32, 68, 106]. Recent years have seen increased global momentum for improving data sharing and, in some cases, a push for open data and lowered barriers to data access [8, 25, 88, 95, 100, 128].¹ However, concerns such as privacy, intellectual property rights, organizational and structural challenges, cultural and social contexts, unjust historical pasts, and potential harms to marginalized communities all make improving data access a highly complex and at times contentious issue [8, 28, 60, 71].

Data sharing challenges are further exacerbated in the African context: Historical and present-day colonial-era practices and ideologies force us to confront a reality characterized by *data colonialism* [29, 36, 79, 115]. Countless data sharing initiatives as well as a growing enthusiasm for data sharing movements can be found throughout the 54 distinct nations in the continent [14, 21, 32, 116]. At the same time, various obstacles limit equitable data sharing practices; entire heterogeneous geographies of people have their data accessed and shared, yet do not reap the same benefits as the data collectors and owners of data infrastructures [40, 87]. These communities can face harm as a result of their data being accessed, shared, and used [84, 93]. At its core, various structural and relational challenges plague the relationships between entities based outside of the continent and the African communities whose data is being accessed. The continent's plural and at times divergent norms, practices, and traditions furthermore complicate the African data access and sharing ecosystem. In sum, data sharing in Africa poses a unique set of challenges that remain understudied.

The African data sharing ecosystem involves numerous stakeholders, including communities, companies, government bodies,

¹While data sharing often focuses on facilitating ease of data transfers between individual scientists, institutes, and/or organizations, truly open data goes beyond that by making data 'freely' available for use to a wider-range of populations, rendering it a public good. In this work, we primarily focus on data sharing practices, although many of the insights we draw also apply to arguments for opening data.



Figure 1: Stakeholders in the African data sharing ecosystem. Those at the top of the iceberg hold significant power and leverage in guiding data sharing practices and policy compared to those in the hidden part of the iceberg. (Image adapted from [41])

policymakers, non-government organizations (NGOs), data collectors, data labellers, data brokers, data subjects, archivists, researchers, and many others [4, 85, 87, 114]. Responsible data sharing requires a harmonious partnership between all stakeholders, both under and on top of the iceberg, as depicted in Figure 1. Consideration of the interests, values, and needs of all — and especially those whose data is being collected — is a vital step in fostering responsible data sharing practices. In many cases, however, we find that more powerful stakeholders wield disproportionate power in driving the design, framing, and enforcement of data sharing policies and practices; their values and interests dictate not only the data sharing landscape, but also narratives around data access.

On the other hand, data sharing in the continent relies heavily on those in the hidden part of the iceberg. Data is often extracted from these same communities, and analyses of the data often have a direct impact on them. Given the significant role of power asymmetries between these groups, data sharing policies and practices that do not acknowledge such imbalances will fail to capture all relevant perspectives. In this work, we examine this dynamic data sharing landscape, its possible benefits and harms, and challenges around data practices. Our work focuses on the following questions: Who benefits from data sharing in Africa? What barriers exist in the data sharing ecosystem, and for whom? How do entrenched legacies of power disparity, colonialisms, ethno-centricisms, slavery, and Anglo- and Euro-centrism present challenges in the data sharing ecosystem? If much of the data sharing practice and initiative is shaped from the perspective of the Global North, how then can we ensure that the narrative for Africa is controlled by Africans? We tackle these questions guided by the experiences and perspectives

of African communities and initiatives, and especially those in the hidden part of the iceberg.

Dominant narratives around data sharing in Africa often focus on a lack of knowledge about the value of data and training, as well as insufficient technological resources. This is often followed by *deficit narratives*,² where discussions regarding data sharing in Africa revolve around issues of hunger, illiteracy, unemployment, and poverty. In anchoring narratives around these concerns — 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. The perception of the dominant (often Western) narratives as the standard comes at the expense of discarding local knowledge and expertise. Furthermore, data sharing issues are discussed as a Global North/Global South concern, ignoring plural norms, practices, and traditions that exist within the continent, which themselves demand a thorough examination.

As we navigate this data sharing landscape, we counter these deficit narratives and clichéd stereotypes with which the African continent is frequently portrayed. In doing so, we hope to highlight overlooked imbalances, inequalities, and injustices in the data sharing ecosystem, and to support the development of policies and programs that will illuminate and mitigate them. We counter these narratives through *storytelling*. We construct fictional personas based on semi-structured interviews with data experts from the continent, as well as our collective expertise in collecting, accessing, and sharing data within the continent. Through these personas, we highlight challenges faced by African communities in accessing and sharing data, as well as the strategies they employ to insist on community-affirming data practices.

The stories also examine the power structures that affect if and how a community benefits from transactions involving the use of their data. By using these personas as touchpoints, we find that three overarching challenges — power imbalances among stakeholders, disinvestment on building trust, and disregard for context and local knowledge — drive inequalities and injustices in the data sharing ecosystem. Through an analysis of these failures, we argue that the policy and practice of data access and sharing must centre the needs of those stakeholders in the hidden part of the iceberg. The values and expertise of communities whose data is being collected and shared need to take precedence in the development of data governance standards. Through our discussions, we advocate for policies and practices that acknowledge and aim to counter power asymmetries and structural obstacles. In presenting our arguments, we also bring to the forefront rich research and scholarship on this topic by African scholars.

2 BACKGROUND AND RELATED WORK

Data sharing and open data movements are gaining momentum globally [26, 32, 64, 100], including in the context of the African continent [13, 25, 34, 34]. Some of this push within the continent comes from funding agencies and government requests for researchers to

²Deficit narratives reduce a group or culture to its “problems,” rather than portraying it with the strengths, creativity, and agency of its people.

archive and share data, including open data portals from countries such as Kenya, Ghana, Nigeria, and South Africa [2, 23, 42, 53]. The call for data sharing is found in various researching fields across the continent, including in the environmental, geospatial, agriculture, health, and biomedical sciences [18, 40, 105, 107]. The intergovernmental Group on Earth Observations (GEO) [9], for example, runs initiatives focused on Africa (AfriGEOSS) with the aim of enhancing earth observation data production, management, and sharing. Similarly, the digitization and sharing of soil data are believed to play a crucial role in a number of fields, including food security, health, hydrological modelling, and climate change [48, 57, 118, 121]. Citing dialogues during a South African Soil Information Workshop, scholars identify the expansion of soil databases and data sharing as important developments for soil science and environmental fields [22, 101]. The biomedical sciences also echo the call for data sharing [3, 27, 40, 50, 86]. In a recent paper, Martin et al. [72] have called for global genomic data sharing and discussed the urgent need for more genetic data studies, while also arguing that African populations are not represented in global genomic databases. Although the authors raise crucial concerns, including the unsuitability of the “European method” to study heterogeneous African populations, they contend that data sharing is vital for a fair representation.

One of the strongest arguments for rethinking how data is shared comes from the Academy of Science of South Africa, which created the African Open Science Platform, a project led by the South African Department of Science and Technology [96]. This project emphasizes the need to increase awareness, accessibility, and visibility of African science and data. It calls for data and science to be as open as possible, with the aim of stimulating interdisciplinary use of data, identifying research areas for further exploration, and averting the duplication of research.

The questions of data sharing and management are becoming ever more pertinent to the continent. In one of the key gatherings in 2018, various stakeholders and government officials from Botswana, Ethiopia, Kenya, Madagascar, South Africa, and Uganda discussed and formulated national policies [122]. With increased demand for statistical data across many African countries, many data organisations have been established to promote data accessibility and open data [8, 24, 34, 100, 128]. For instance, the African Development Bank Group developed a statistical open data portal known as Africa Information Highway (AIH) portal [46]. AIH links all African countries with the aim of increasing public access to statistics across the continent, while also helping these countries improve data management, data quality, and sharing/dissemination. Similarly, Kenya Open Data [86] is a portal that provides public access to several governmental datasets curated in different categories at no cost. This platform aims to support the government’s commitment to transparency and data sharing between government offices and various stakeholders.

Despite extensive calls for data sharing and open data, backed by various initiatives and organizations, data access remains relatively limited within the continent [34, 43, 113, 119]. Bangani and Moyo [18], for example, compare data sharing practices among researchers in South Africa to researchers in countries such as France, the United Kingdom, and Turkey. The authors state that most researchers in South Africa are not interested in sharing their data and prefer to use data generated by others; only 19.4% of South African

researchers indicated that they currently share their research data with others. As reasons for their reluctance, these South African researchers cited worries ranging from unethical use or misuse to lack of resources. In extended fieldwork with biochemistry laboratories in sub-Saharan Africa, Bezuidenhout and Rappert [28] find that scientists opt for closure over openness with regard to their data. The reasons put forward include inequalities in research environments which, if not examined critically, perpetuate disparities. Furthermore, an overwhelming number of researchers in Africa lack the funding necessary to make their data open to the wider community [26]. Bezuidenhout and Chakauya [26] report a lack of financial investment in physical infrastructures could make a researcher’s data unavailable, especially for critical clinical research. Barnes et al. [19] agree that funding is a panacea for doing quality research and promoting data openness in low-resource settings. We note, however, that enthusiasm levels for sharing data vary from field to field [14, 43, 109, 119].

Obstacles to data sharing are not only financial or infrastructural, but also extend into the political sphere. Bezuidenhout and Rappert [28] raise a fundamental question regarding open data and open science. Their study shows that data sharing and open data practices are informed by Western perspectives, and are driven by the contextual issues of Western researchers. Given that different circumstances will give rise to different issues, they go on to ask whether the value attached to data sharing is a specifically Western one. The question becomes: 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. Yilma [126], for example, has proposed the absence of a strong political will as a major barrier to data sharing in Africa. Likening data sharing practices to aid programs, Yilma states, “any type of aid program that does not lead eventually to self-sufficiency is actually destructive, just like welfare, and that is what has happened with many aid programs in Africa” [126].

3 STORYTELLING AS A METHODOLOGY

“Stories are just data with soul” – Brene [31].

In this section, we illustrate the importance and relevance of storytelling to our work. We first trace its historical significance to the African continent before discussing our methodology for creating personas aimed at presenting counternarratives.

3.1 Storytelling in Africa

Oral culture and the African continent are profoundly linked in most spheres, from poetry and dance to philosophical traditions [92, 99, 111, 120]. Many philosophical ideas have deep roots in the African continent [51, 76, 81, 103]. Ancient Egypt, for example, marks the oldest known philosophical tradition [54, 67, 97]. Furthermore, in pre-colonial Africa philosophical practices among indigenous peoples took the form of oral traditions [6, 55, 108]. Emphasizing the distinctness of oral traditions to the African continent, Okot p’Bitek [102] contends that oral practices are much more immediate and powerful. In *Oral Traditions as Philosophy: Okot p’Bitek’s Legacy For African Philosophy*, Imbo [55] further

demonstrates the unique significance of oral traditions in African philosophy.

Storytelling is one important method for indigenous research, and a tool to legally claim the right to local voices through which narratives are validated [39, 47, 56, 73]. Making local communities the focal perspective of shared knowledge through storytelling counters histories of colonialism, knowledge produced for colonial regimes, and the power dynamics silencing indigenous expertise [10, 17, 77, 110]. In South Africa, for example, the *Truth and Reconciliation Commission* (TRC) incorporated storytelling – songs, poems, and oral literature – into the official legal process of speaking aloud marginalized experiences of Apartheid [94]. In Uganda, the National Education Ministry mandates storytelling as part of the required national curriculum. The Ugandan Education Ministry’s curriculum reform since independence has likewise re-introduced storytelling by law. These reforms are focused on re-instituting local knowledge and teaching methods, as well as countering the influences of former colonial powers [15, 83, 90, 112]. The African StoryBook Project³, a repository of stories in local African languages, illustrates the value of both storytelling as a method of knowledge creation and data sharing practices driven by community needs. This open source repository of digital books written in local African languages allows the re-purposing, scaling, and sharing of content from the platform, helping the Ugandan Education system to make the switch to teaching in local languages⁴.

African researchers and scientists have called for the international scientific community to embrace indigenous research [20, 33, 38]. The call acknowledges how research design (e.g. how “validity” is determined) could reflect Western values, and thus, co-opt, erase, and/or circumvent legitimate means of expressing knowledge. Problematically, many Global North research methods characterize local ways of creating and validating knowledge as *scientifically invalid* [44, 63, 125]. Computational journal and conference reviewers also insist on Global South authors contextualizing research, including methods, in relation to the Global North. We argue this demand erases and marginalizes local expertise.

3.2 Storytelling of African Data

Inspired by the role of storytelling in the African context, we illuminate challenges faced when accessing and sharing data in the African continent from African perspectives. We do so through personas created to portray recurring challenges we encountered throughout our interviews, as well as our experiences of collecting, accessing, sharing, and using data generated in the continent.

We constructed the personas through an iterative process based on semi-structured interviews with ten African data experts. Our interviewees included individuals working in universities, governments, startups, and think tanks, and we focused on data collected and shared for research or policy purposes. We first developed open-ended questions, guided by our experiences observing and

³<https://www.africanstorybook.org/>

⁴Although the African StoryBook Project can be seen as a powerful example, it is also important to acknowledge that limitations and challenges remain for the Ugandan Education system. For instance, the harms of colonial regimes are not extinguished immediately upon instituting policy reforms. The dynamics of the global political economy also continue to pressure national education systems to emphasize teaching and establishing grading or progress standards for students that are measured by English language competency, e.g. standardized tests

actuating data sharing practices on the continent. The interviews were recorded with permission from the interviewees, after which we manually transcribed the audio recordings. Next, we identified and abstracted central themes, challenges, and concerns expressed by the interviewees. Any information that could be used to identify our interviewees, institutions, government bodies, or communities was removed. In addition, we only took a particular challenge as the inspiration for the creation of a persona if that challenge appeared in at least three interviews. These decisions allowed us to preserve the privacy of both the individuals we interviewed and the communities of concern, while still creating space for discussing common data sharing practices that are not affirming of the needs, values, and concerns of those on the bottom of the iceberg. In response to these privacy concerns, we did not incorporate any direct quotations from our interviews into the personas, but we plan to include data of this kind in subsequent publications.

Each persona consists of two to three core challenges. Over the course of multiple meetings, we discussed and debated the efficacy of each persona in portraying the challenges we would like to elucidate. In our discussions, we primarily focused on the following key questions:

- (1) Do the personas elucidate the intended data sharing challenges, and if not, why not? How can we better bring these challenges to the surface?
- (2) Which data sharing challenges are more “familiar” to a diverse audience, including many researchers from the Global North? In what ways should we amend the personas to hone in on what challenges may be unique to the continent?

This process allowed the personas to evolve and solidify over time. We rebalanced the emphasis on various themes in the personas at times by adding details from more interviews or removing details from interviews that obfuscated the key challenges we would like to highlight. In this paper, we present a sample of three personas that illustrate the three key overarching challenges we found in our work. Additional personas and details about this project are shared on our project page.⁵

4 PERSONAS

“Proverbs are horses on which words ride, when words are lost [i.e. when the truth of a matter is elusive], it is proverbs we use to search for them [i.e. proverbs help us to solve an intractable problem]” – Agbájé [11].

The personas we present below present recurring themes we encountered in our interviews and subsequent discussions. These stories are not meant to reflect the experiences of any single individual, community, company, or country, but rather to gather evidence and shed light on recurring barriers. Since these stories are based on aggregate information, they call attention to common social issues while respecting sensitive dynamics – such as concerns about privacy and retaliation – that may make interviewees, and indeed many across the continent, vulnerable when sharing these experiences. As such, these stories convey and conceal truths at the same time.

⁵<https://www.md4sg.com/workinggroups/GlobalSouthData.html>

On Good and Harm

Themes: power imbalance, data colonialism

Context:

DrinkUS West is a Europe-based non-government organization (NGO) working on sustainable access to water and health. DrinkUS hopes to support people in Buranda – home to the yetet'ebek'e community – by improving access to clean, potable water. It also plans to deploy and test new water accessibility technology and online monitoring of resources.

Data needs:

To provide adequate services and assess its impact, DrinkUS seeks data about people's day-to-day water use habits and experiences through surveys. In the survey, DrinkUS researchers ask individual respondents about their demographics (age, gender, income, marital status, number of household members and children), discretionary spending habits (ownership of computers, mobile phones, leisurely activities), and water use habits (such as how water and utilities are accessed, and the respondent's knowledge of potable water accessibility). Additionally, DrinkUS seeks country-wide data on demographic patterns as well as outcomes related to health, education, and poverty levels over the past decade. It gains access to this data from the Buranda Health Ministry.

By using these two data sources, DrinkUS aims to estimate demand for potable water, determine prices, and predict the impact such a service may have on the health and economic welfare of the community. Additionally, the NGO assesses technologies that could be used to monitor and manage water access. As the NGO is keen to engage survey participants in future opportunities, it collects contact information such as the names, addresses, and phone numbers of participants. DrinkUS incentivizes each respondent by paying the equivalent of \$5 U.S. dollars in Kwacha, the local currency, for completing the estimated 30-minute long survey to improve survey completion rate. From this data, DrinkUS researchers compile and infer which members lack access to potable water, what community members can afford to pay, and who is most at risk for water-related health impacts and/or disease. They combine these results with data obtained from the Buranda Health Ministry to gain deeper insights into the impact of access to potable water. This work is later published as a research paper at a top-tier international conference on data science for social good.

Data sharing challenges:

While employees at DrinkUS meticulously checked their survey design, they did not seek adequate feedback from the yetet'ebek'e community beyond a few initial meetings. The yetet'ebek'e community is not treated as a mutual partner in the broader project design. In fact, there is an implicit assumption that improving water access for the community is an adequate benefit, and that DrinkUS is therefore justified in collecting and owning the survey data. Further, since DrinkUS is a world-renowned NGO working in partnership with the Burandan government, it wields significant power over members of the yetet'ebek'e community. Even the pay incentives, which may be negligible for the NGO, are significant due to currency exchange rates, further highlighting a power imbalance.

Given that the NGO proceeded without consulting the community, it likely failed to grasp all of the central issues and potential

harms from the perspective of the community. 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, put the community at risk, as many belong to marginalized ethnic groups. Such findings, even shared in aggregate, can reveal information on the community as a whole. While notions of privacy often focus on the individual [52, 65], 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 [45, 89].

In addition, given the partnership with the Burandan government and poorly defined protocols around data sharing and use, survey participant data was shared with government officials, including data about household composition and reported or predicted health challenges linked to potable water access. While the researchers vetted their questions for survey design validity, they did not seek active involvement from the yetet'ebek'e community to discuss whom the data can be shared with and for what purposes. DrinkUS also failed to solicit feedback on the broader project design, data ownership and its use beyond original scope of work, and research procedures. This disempowerment of local communities in the decision-making process introduced trust issues both with the NGO as well as with the Burandan government, which has been known to engage in open data sharing practices with large NGOs and institutions. Local communities, on the other hand, cannot access their own data with the same ease.

The Journey of African Scholars

Themes: trust, awareness

Context:

A doctoral candidate from Bozatta researching in Nova Africa wants to collect soil samples from several Nova Africa provinces for her research. Her study aims to investigate the fertility of soil samples, which will inform key stakeholders, such as the Government and NGOs, to find the best approaches to assisting farmers.

In another case study, a doctoral candidate is researching the prevalence of gender-based violence in Nova Africa. The study is targeted towards raising awareness of gender-based violence prevalence, which remains a profound and widespread problem in Nova Africa.

Data needs:

Case 1: Soil and apartheid

The researcher has outlined critical steps to conducting the soil fertility test, using techniques such as sampling, analysis, interpretation, and recommendation. All of these methods are essential for estimating soil fertility. Enthusiastic about undertaking her research, the researcher aims to collect these samples to determine soil fertility, nutrient deficiencies, and recommendations, such as selecting the most favourable fertilizer plan to increase farmers' productivity. The information gathered from the study will be used to create an effective fertility management program and provide findings to the public through her dissertation and institutional public databases.

Case 2: Gender-based violence

The researcher aims to collect this data and create a data platform to present gender-based violence cases within provinces in Nova Africa. The tool will aid data visualisation of these cases and will be available to the broader public. The researcher understands that her work is focused on a pressing issue: Gender-based violence remains a persistent problem, requiring attention from multi-faceted responses and commitments from the government, civil society, and other citizens. In particular, the government does not keep track of data on the prevalence of gender-based violence, despite desire to work towards protections for the vulnerable. The researcher's work aims to narrow this data gap.

Data sharing challenges:

In *Case 1*, in an attempt to collect soil samples, the researcher 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 colleagues that farmers have collected such data for their own internal purposes, and use it to organize within their communities. The researcher attempts to instead obtain this data, assuming that the farmers' reluctance may be partially due to replication of efforts. She eventually learns that the farmers are reluctant to share any information due to fear that the government might want to claim ownership of their lands. Collection of such data is a challenge for local researchers as well. This issue is exacerbated by the history of a former apartheid regime, under whose reign land grabs were widespread, impacting Black Nova Africans primarily. Since sharing data can lead to forfeiture of land rights, trust is a central issue for the farmers in deciding how much information they share and with whom.

In *Case 2*, during the process of conducting the research, the doctoral candidate experiences a few unexpected encounters and shares them with some colleagues from Nova Africa. She tells them that many participants became emotional when sharing their experiences, and asked that the session be discontinued. One participant even questioned why the interview was so intrusive, and decided not to share further information. Her officemate tells her that in Nova Africa, it is challenging to convince participants to release information, especially as a foreigner. There is a perceived notion that foreigners are here to steal from them, whether that be data or resources. This, she learns, is in part reflective of and caused by divisions across communities in Africa due to colonialism. Another colleague advises her that some participants might seek secluded venues that are secure for sharing experiences in most rural areas, but that building trust is crucial, especially as an outsider of this community.

Young researchers working across different African countries may encounter challenges when trying to collect, access, and share data. This challenge is particularly stark when the researcher is not a local to the region or community in which the data is being collected. Differences in languages and cultures, as well as imbalances in trust and power, breed suspicion about motives, impeding data access and sharing. With more than 1500 languages spoken across the African continent, locals residing in rural areas are more likely

to communicate in their dialect, with potentially little or no understanding of national languages. This concern may, at times, stand in contrast to the fact that there is a pressing need on high-quality data to understand the prevalence and impact of issues such as gender-based violence, among numerous other forms of injustices and discrimination.

Discussions on data sharing in the African context often treat the continent as a monolith and frame problems as a Global North/Global South issue. However, the complex data sharing ecosystem can lead to challenges even for those based in the continent. We further find that many issues with power imbalances and trust can be traced back to the impact of entities based outside of the continent. We see, for instance, that certain hesitations with sharing data are rooted in colonial-era extracting practices whose impact can be felt to this day. Numerous efforts to mitigate inequalities in data access and sharing in the continent, and especially those driven by initiatives outside of the continent, may not be attuned to the complex data landscape resulting from a diversity of needs, priorities, and experiences of those in the continent.

Livestocks and Livelihoods

Themes: data infrastructure, awareness

Context:

A researcher in Wolonda receives a two-year grant from a large foundation from the Global North. The research project outlines an ambitious goal of ushering in prosperity for the livestock farmers by making the mainstream markets work for them.

Data needs:

The researcher outlined his intentions to collect various types of data, including genomic data of crops and livestock and enterprise data for the farm. The genomic dataset would help determine if the right crops were being planted and the right animal breeds 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 endeavor, with clear data collection and analysis objectives. Yet no protocols for data storage and scope of use were defined, and there were no clear guidelines on whether the data can be shared with third parties. Furthermore, scant information was provided as to if and how the collected data would be used beyond the time period of the donor-funded project.

As expected, the researcher and his team diligently collected the genomic data as well as the farm data from a cluster of villages in several counties of Southern Wolonda. The data collection process was, however, 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.

Data sharing challenges:

Encouraged by this newly collected and valuable dataset, a research team 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 dataset which was collected as part of the original research project. In addition, due to infrastructure challenges, the data is now housed in a data center 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 Wolonda. Even though the trust established with the farmers in Southern Wolonda has eroded due to 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.

Foundations in the Global North may be eager to fund data collection projects in an attempt to create an oasis in regions perceived to be data-starved – to bridge data gaps and inequalities that appear to be merely inefficiencies. Yet, as is the case in this scenario, the lack of forethought into the ethical use of data – not only during the lifetime of the project, but also thereafter – can create irreparable harms to communities' well-being and make it difficult to build partnerships based on respect and mutual trust. This above case may alienate the farmers, precisely the people whom this project was ostensibly seeking to support.

5 DATA SHARING CHALLENGES

As the personas in the last section show, obstacles, issues, and challenges of data sharing concerning the African continent are multifaceted. Questions such as, 'Is data sharing beneficial? Valuable?' need to be contextualized by subcommunity: We need to also ask 'for whom?' Throughout this work, we have examined such questions in a manner that is informed by the perspective of the stakeholders at the bottom of the data ecology chain – the hidden part of the iceberg. There are numerous causes for challenges in this ecosystem. In this section, we focus on three broad and inter-related challenges that surfaced from our interviews and research. In particular, we explore: power asymmetries, issues of trust, and the need for contextual knowledge.

5.1 Power asymmetries

Historically, traditional colonial powers sought unilateral domination over colonized people through control of socio-economic affairs and the reinvention of social orders for their own benefit [98]. In the current day-and-age, when data and digital technologies are powerful instruments, colonial-era oppression has been reincarnated in various data practices, including data collection, sharing, and analysis [29, 62]. The search for data accumulation, especially with regard to Western technology monopolies – both the scale as well as the manner in which it is being collected – have raised questions regarding unprecedented wealth accumulation and power struggles. Currently, a significant proportion of Africa's digital infrastructure is controlled by Western technology powers such as Amazon, Google, Facebook, and Uber. Traditional colonial powers pursued colonial invasion through justifications, such as "educating the uneducated." Data accumulation processes are accompanied by similar colonial rhetoric, such as "liberating the bottom billion," "helping the unbanked," "connecting the unconnected," and using data to "leapfrog poverty." However, this rhetoric

may not only preserves historical coloniality [80] dressed in data, but also perpetuates deficit narratives. These objections have been articulated by scholars and technologists within the continent. For instance, Kimani [62] writes:

"I find it hard to reconcile a group of American corporations, far removed from the realities of Africans, machinating a grand plan on how to save the unbanked women of Africa. Especially when you consider their recent history of data privacy breaches (Facebook) and worker exploitation (Uber)." [62]

Colonial legacies and power imbalances embedded in data practices may appear less obvious and more nuanced within scientific research settings [16]. Within the context of global health research, there has been growing concern about power imbalances in authorship, which has negatively affected how research undertaken by local researchers in low-income and middle-income countries (especially in Africa) is perceived [1]. Academics in the Global North are bestowed with the power to define what constitutes "legitimate knowledge," "good research," "standard method," or a relevant and worthwhile problem. As a result, African researchers are left little room, if any at all, to compete at the global stage, even in matters concerning the African continent. Following extensive analysis of interviews with senior university research managers in Zimbabwe, and on a public roundtable on Structural Inequalities in Global Academic Publishing, Jeater [59] finds that "When we ask who gets to represent the 'African perspective,' we find it is decreasingly un-African." Resource inequalities, hegemonic academic standards that undervalue Southern research traditions, and the unilateral power Northerners hold to validate research, all contribute to structural obstacles that amount to systemic exclusion of African scholarship from global health research. In a data ecosystem built on such firm yet invisible power asymmetries, stakeholders already in a position of power not only benefit the most, but also make data accessibility inequitable.

Imbalances in authorship and power asymmetries constitute a continuation of the colonial project in global health research, creating fundamental trust issues concerning data sharing. Underneath power asymmetries in global health research partnerships between researchers from the North and South, lies what Abimbola [1] calls "the foreign gaze." In his analysis of this concept, Abimbola asks questions such as "who we are as authors, who we imagine we write for (i.e., gaze), and the position or standpoint from which we write (i.e., pose)." Furthermore, power asymmetries which occupy an important space around trust and data sharing often operate in invisible ways and take many forms.

In a similar study that explores power imbalances between the global North and South, using Zambia as a case study, Walsh et al. [124] report that power imbalances and inequalities manifest at all stages of research. This includes everything from funding to agenda setting, data collection, analysis, interpretation, and reporting of results. Looking at the underfunded Zambian health research, where up to 90% of the funding for health research comes from external funders [124], the bargaining power rests with the funders with little room for negotiations for Zambian scholars. Power asymmetries are also observed in the personas above. In "On Good and Harm", we find the European NGO at the top of the power hierarchy making

key decisions regarding the data concerning yetet'ebek'e communities. Similarly, in the “*Livestocks and Livelihoods*” persona, the foundation from the Global North which provided grants dedicated to research in Wolonda holds much greater authority and power compared to the villagers who are the source of the data.

Power asymmetries, historically inherited from the colonial era, often get carried over into data practices and manifest themselves in various forms, from imbalanced authorship to uneven bargaining powers that come with funding. Having said that, power asymmetries are not limited to historical contexts only. Within a given research project, for example, one can observe that power asymmetries exist between project managers and data analysts; data analysts and data collectors; data collectors and research participants. All these factors are contingent on various caveats and implications for trust impacting data processes from data quality to data sharing. Examining data production and consumption process in the context of Malawian demographic survey, Biruk [30] makes these power asymmetries, hierarchies and structural inequalities visible. Although obscured by partnership rhetoric, Malawian demographics mapping in fact embodies structures with unequal division of labor. Biruk explains:

“[B]eing on the ground in the field has the largest effect on data but—from the perspective of researchers—the activities of fieldworkers are framed as menial labor performed by easily replaceable and interchangeable individuals. The local expertise they offer, then, is not in designing research or writing proposals but comes as an additive to a project conceived in a distant office. These hierarchies are embedded in political-economic structures that privilege the knowledge work that is the purview of Western academic researchers over the so-called unskilled labor performed by field workers. Meanwhile, Malawian research collaborators occupy a middle space that is both constructed by and fraught with power and economic inequalities.” [30]

5.2 Trust

“*Data often move at the speed of trust.*” – Hamilton and Hopkins [104]

Sharing data between different stakeholders hinges on trust. Trust is the fundamental component of all relationships in a data sharing ecosystem. While trust, or lack thereof, has been identified as a key challenge that hampers data sharing [7, 37, 58, 109], there remains much to be examined about the role of trust and how it manifests in relationships between various stakeholders in the African data sharing ecosystem. Data sharing practices which operate in the absence of knowledge of local norms and contexts contribute – albeit indirectly – to the erosion of trust among stakeholders in the data sharing ecosystem. Initiatives coming from outside, with their own assumptions, interests, and objectives, tend to be met with suspicion by local communities. The persona on “*Soil and Apartheid*” captures this in a stark manner; due to the apartheid regime, the doctoral researcher finds that Black farmers in Nova Africa suffered unjust land grabs. Such historical injustice plays out in the farmers’ reluctance to share soil data due to lack of trust.

Resource inequalities and colonial oppressive histories instill deep mistrust towards open data and data sharing initiatives. Although African researchers are generally supportive of data sharing, they are considerably less enthusiastic about open data expressing concerns that open data compromises national ownership and reopens the gates for “parachute-research” (i.e., Global Northern researchers absconding with data to their home countries). Such concerns are not unwarranted. In fact, the findings from a recent study from Mbaye et al. [75] affirm this fear. Mbaye et al. [75] performed a systematic review examining African author proportions in the biomedical literature published between 1980 - 2016, in which research was originally done in Africa. The authors found that African researchers are significantly under-represented in the global health community, even when the data originates from Africa. A common threat is parachute-research, in which non-African researchers benefit from data sharing and open data, are afforded the opportunity to narrate African stories (in some cases also contributing to deficit narratives), and publish scientific work using African generated data available through open access initiatives – all while ignoring the contributions of African communities and scholars [66, 74, 82, 127]. Recent work within the medical sciences published on the Ebola outbreak in the Democratic Republic of the Congo can be cited as a prime example in this regard.

Ideal data sharing initiatives, policies, and principles weigh the benefits and potential risks, and strive to find a reasonable balance. Benefits and risks also vary depending on the types of data being shared. For example, the use of genomics data poses the potential for far more detrimental risk to the individual providing the data, to the researcher, to the institution, and to the community compared to soil data. The issues of trust as a challenge to data sharing, likewise, vary in degree depending on the data in question. Trust is a relatively significant challenge to sharing sensitive biomedical or health data and less significant when it comes to, for example, environmental sciences data. Noting this concern, Walport and Brest [123] emphasize, “people often agree to provide sensitive data because they trust the researcher and believe the researcher would not use the data in a way that would be harmful towards them. However, there is a concern that the trust may not carry over when the data are shared”.

5.3 Contexts and local knowledge

Contexts are a crucial element to making sense of data and data sharing. Yet data – within data science and machine learning, particularly – are often stripped away of contexts [35, 70, 78]. In the process of data cleaning, for example, information that provides contexts about the specific background from which data are collected and how datasets are structured can be lost [5]. The importance of context for datasets has been explored by scholars, such as Loukissas [69], according to whom, we should shift into thinking in terms of *data settings* instead of *datasets*. Contexts are crucial to understanding data fully; data sharing practices that discard contexts risk becoming irrelevant and potentially harmful to local communities, as we see in each of the above personas.

Consequently, the context of the data itself – which provides a complete image – and awareness of local norms, cultures, and histories constitute crucial elements in a responsible data sharing

practice. Thus, for data sharing practices to benefit the underserved, such groups' welfare and interests need to be placed at the center stage. However, our interviewees nearly unanimously agree that there remains a lot to be done to acknowledge and incorporate the interest, norms, and context of these below the iceberg in Figure 1. In fact, oftentimes, certain groups such as data subjects are hardly recognized as stakeholders at all [12, 49, 61].

We see various levels of disregard for context and local norms displayed in all the personas portrayed in the previous section. In "*On Good and Harm*", the failure to ground data collection in communal understanding of privacy, for example, resulted in the NGO exposing yetet'ebek'e community to various risks in the process of data sharing. Lack of common language and understanding of local norms creates a challenge for intra-continental data practice, as shown in "*The Journey of African Scholars*." In both cases, the doctoral candidate faced difficulties understanding the local norms, which played a role in the community's lack of trust towards her. Similarly, in the "*Livestocks and Livelihoods*" persona, we observe that the funding foundation (from the Global North) failed to consider the context in which the data was collected. In summary, data sharing calls that are not aware of local norms, contexts, and culture, when imposed from the outside, constitutes a form of Western-centrism and colonialism.

6 DISCUSSION

Returning to the question "is data sharing good/beneficial?" we argue that responsible data sharing practices must, first and foremost, benefit local communities and experts, with a focus on those at the bottom of the iceberg. In recent years, the African continent as a whole has been considered a frontier opportunity for building data collection infrastructures. The enthusiasm around data sharing, and especially in machine learning or data science for development/social good settings, has ranged from tempered discussions around new research avenues to proclamations that "the AI invasion is coming to Africa (and it's a good thing)" [91]. In this work, we echo previous discussions that this can lead to data colonialism and significant, irreparable harm to communities. As we learned from the rich body of previous works, our experiences, interviews and the personas, data sharing practices are divergent, ad hoc, at times contradictory, and/or violate community values for data use. As machine learning and data science move to focus on the Global South and especially the African continent, the need to understand what challenges exist in data sharing, and how we can improve data practices become more pressing.

Despite significant variation among the different countries on the continent, many countries are now producing high-quality data in a more reliable and regular manner through statistical offices, national efforts, and involvement of other public agencies. Likewise, non-public sectors actors — such as private-sector companies, academic institutions, civil society organisations, development agencies, and individuals or communities — are now encouraged and incentivized to be involved in open data management. The future of open data management and data sharing and their contribution to the advancement of science and technology in Africa will continue to increase, despite the slow pace caused by the lack

of funding, redundant policy frameworks, and limited infrastructures. The unique African landscape, and especially the existing challenges and how they can be addressed, will continue to play a big part in African participation in open science and open data global projects [26].

As explained by Bangani and Moyo [18], to understand the sharing practices of researchers, it is essential first to know where the data comes from and the lives, cultures, and communities it represents. Bangani and Moyo [18] also conclude that most of the emerging researchers in South Africa would prefer to use data from other sources and are not interested in sharing their own data. Most of these emerging researchers indicated a lack of resources, misuse of data, and trust issues as the significant challenges they face in sharing their data. Some of them indicated that they would prefer to only share their data with collaborators within their universities and not with collaborators outside their universities. For many of the researchers in this context, the labour, effort, and time it takes to collect data is unmatched by the benefits (often little if any at all) they might receive from data sharing. Subsequently, as well as encouraging researchers to share their data, there also needs to be actionable benefits and rewards, as well as policies that protect African researchers and communities from practices like parachute research.

Open data imposes risks to African communities — in terms of both stripping data away from contexts and reduced benefits to communities where data is sourced. It may present a possibility for Global North researchers to publish research using data from Global South with neither acknowledgements to data collectors or data subjects nor an understanding of the data setting. Consequently, there should be less pressure to open up data and more incentives to share data with the necessary conditions and support for infrastructures. The necessary conditions could include data sharing guidelines that give adequate time to African researchers to make use of data before wider sharing [117].

There are various ways towards equitable data sharing that benefit of local communities. As we have seen in the data sharing challenges section (Section 5), responsible data sharing not only requires knowledge of the context of data, it also needs to center the interests and norms of communities where data comes from. Data sharing calls that are not aware of local norms, contexts, and culture, when imposed from the outside, constitute a form of Western-centrism and colonialism. For those coming from outside to participate in data sharing activities, learning "the ways of the people" without infiltrating their cultures need to be considered a vital part of the process. Providing sufficient support to local researchers with the aim of facilitating local-led data sharing practices can be one of the initial steps.

Trust is critical during data sharing. Trust is built if those involved know each other. The reason for sharing is explicitly stated to all the stakeholders involved. Each stakeholder performs their responsibility as agreed, and all stakeholders' expectations are defined, understood, and respected. Supposing all stakeholders agree to these conditions, the likelihood of further collaboration and the possibility of fair sharing increases. It is recommended that the process of data sharing should be transparent. Also, researchers based in the continent are encouraged to build a trustworthy collaboration with funders and other stakeholders. Introducing proper

and adequate legal frameworks, documentation, and support for data sharing will increase trust among stakeholders. Due to uncertainties and lack of legal frameworks, several sectors in Africa are unsure of data sharing norms and practices. They are allowed to share their data, what type of data can be shared, who, and under what conditions. Such frameworks and documentation can help clarify the benefits, expected norms, and equitable data sharing practices. Such can help build trust and encourage researchers to share their data in a manner that acknowledges their labour and protects their ideas and innovations.

Equitable data sharing is also a matter of training the next generation of data scientists. It is imperative to include African data challenges into the data science curricula since these issues are very different from other parts of the world. Aspiring data scientists must be aware of data sources, and historical records about communities where data are collected and the harm these collected data may pose.

Many of the challenges we have discussed here – such as data sharing in the absence of awareness of local contexts and norms as well as the issue of trust – are daunting, and lack simple answers. Nonetheless, as we have sketched in the above sections, they are challenges that will need to be overcome to reduce community harm. The much larger and nuanced issues of data sharing, such as power asymmetry and underlying coloniality, Euro and Anglo-centrism, white supremacy, are obstacles that need deeper structural and social changes. They present challenges with no clear solutions. Understanding and acknowledging their invisible operations is a vital part of the process, and may point to how power asymmetries can be re-balanced. In this work, we have brought to light these invisible forces. While our work focuses on data access and data sharing, we note, also that many of the issues discussed may impact processes across the data generation, collection, labeling, analyses, implementation, and use pipeline. Each of these, we believe, warrants further investigation.

Last, beyond shedding light on the complex data sharing and data practice challenges in the continent, our work also uplifts research on this topic done by Africans for Africans. This practice of centering African scholarship itself challenges negative stereotypes of the continent as “uneducated population.” There is a vast body of work examining the rapidly evolving data practices of the African continent and its relationship with the global data ecosystem.⁶ In the same way that data sharing challenges cannot be overcome without understanding local contexts, meaningful solutions will not be reached without deep engagement with African scholarship.

7 ACKNOWLEDGEMENTS

Abeba Birhane was supported, in part, by Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero, the Science Foundation Ireland Research Centre for Software, Ireland (www.lero.ie). The authors would like to thank our interviewees, members of the Data Science Africa initiative, workshop and conference participants at the Machine Learning for Development (ML4D) workshop, the Privacy Law Scholars Conference (PLSC), the Contested Data Academic Workshop at

Data and Society, the Practical Machine learning for Developing Countries (PML4DC) workshop, as well as members of the MD4SG Data Economies and Data Governance working group for invaluable feedback throughout the development of this work. We would also like to thank Maria Tatar, Thomas Laurent, and anonymous reviewers for the same.

REFERENCES

- [1] Seye Abimbola. 2019. The foreign gaze: authorship in academic global health. *BMJ Global Health* 4, 5 (2019), 1–5.
- [2] Sulemana Abubakar. 2019. *Towards a Decade of Open Government Data in Africa: A Fit Viability Case Analysis of Ghana*. Ph.D. Dissertation. University Of Ghana.
- [3] Sally Adebamowo, Veronica Francis, Ernest Tambo, Seybou Diallo, Guida Landouré, Victoria Nembaware, Eileen Dareng, Babu Muhamed, Michael Odutola, Teniola Akeredolu, et al. 2018. Implementation of genomics research in Africa: challenges and recommendations. *Global Health Action* 11, 1 (2018), 1419033.
- [4] AA Adekunle, Jim Ellis-Jones, Igbekele Ajibefun, Rose Nyikal, Solomon Bangali, Oluwole Fatunbi, and Alain Angé. 2013. *Agricultural innovation in sub-Saharan Africa: Experiences from multiple stakeholder approaches*. Forum for Agricultural Research in Africa (FARA), Accra, Ghana.
- [5] Aremu Adeola. 2020. Lost in translation: Why Google translate often gets Yorùbá – and other languages – wrong. <https://globalvoices.org/2020/03/03/lost-in-translation-why-google-translate-often-gets-yoruba-and-other-languages-wrong/>. Accessed: 30 September 2020.
- [6] Michael Adeyemi and Augustus Adeyinka. 2003. The principles and content of African traditional education. *Educational Philosophy and Theory* 35, 4 (2003), 425–440.
- [7] Afe Adogame. 2014. Reinventing Africa? The negotiation of ethnic identities in the new African religious diaspora. In *Religion, ethnicity and transnational migration between West Africa and Europe*. Brill, Leiden, The Netherlands, 12–36.
- [8] Eric Afful-Dadzie and Anthony Afful-Dadzie. 2017. Open Government Data in Africa: A preference elicitation analysis of media practitioners. *Government Information Quarterly* 34, 2 (2017), 244–255.
- [9] AfriGEOSS. [n.d.]. Enhancing Africa’s capabilities for producing, managing and using Earth observation data and information. (n.d.). <http://www.unspider.org/sites/default/files/Africa%20Book%20AfriGEOSS%20article.pdf>
- [10] Odero Agan. 2006. Storytelling as a means of disseminating knowledge in museums: the example of Sigana Moto Moto. *Museum International* 58, 1-2 (2006), 76–82.
- [11] Bôdè Agbájé. 2002. Proverbs: A strategy for resolving conflict in Yoruba’ society. *Journal of African Cultural Studies* 15, 2 (2002), 237–243.
- [12] Kanayo Alaneme and Anita Okotete. 2018. Critical evaluation of seismic activities in Africa and curtailment policies—a review. *Geo-environmental Disasters* 5, 1 (2018), 24.
- [13] Evelyn Anane-Sarpong, Tenzin Wangmo, and Marcel Tanner. 2020. Ethical principles for promoting health research data sharing with sub-Saharan Africa. *Developing World Bioethics* 20, 2 (2020), 86–95.
- [14] Evelyn Anane-Sarpong, Tenzin Wangmo, Claire Leonie Ward, Osman Sankoh, Marcel Tanner, and Bernice Simone Elger. 2018. “You cannot collect data using your own resources and put it on open access”: Perspectives from Africa about public health data-sharing. *Developing World Bioethics* 18, 4 (2018), 394–405.
- [15] Ketty Anyeko and Tamara Shaya Hoffmann. 2020. Storytelling and Peacebuilding: Lessons from Northern Uganda. In *Peacebuilding and the Arts*. Springer, 235–251.
- [16] Bill Ashcroft, Gareth Griffiths, and Helen Tiffin. 2013. *Post-colonial studies: The key concepts*. Routledge, London.
- [17] Johnson Oluwole Ayodele. 2020. Colonialism and victimization narratives in the context of Africa’s development. In *Global Perspectives on Victimization Analysis and Prevention*. IGI Global, USA, 54–74.
- [18] Siviwe Bangani and Mathew Moyo. 2019. Data Sharing Practices among researchers at South African universities. *Data Science Journal* 18, 1 (2019), 28.
- [19] Karen Barnes, Arturo Canario, Sunil Vernekar, Shivaprasad Goudar, Roberto Espinal, Laura Merson, and Yeong Cheah. 2019. Equitable data sharing: challenges and suggestions for ways forward. *Wellcome Open Research* 4 (2019), 172.
- [20] Cyndy Baskin. 2005. Storytelling circles: Reflections of aboriginal protocols in research. *Canadian social work review/revue canadienne de service social* 22, 2 (2005), 171–187.
- [21] Michael Batuo, Kupukile Mlambo, and Simplice Asongu. 2018. Linkages between financial development, financial instability, financial liberalisation and economic growth in Africa. *Research in International Business and Finance* 45 (2018), 168–179.
- [22] Marye Belete, Jinsong Deng, Ghali Abubakar, Menberu Teshome, Ke Wang, Muluneh Woldetsadik, Enyan Zhu, Alexis Comber, and Adam Gudo. 2020. Partitioning the impacts of land use/land cover change and climate variability on water supply over the source region of the Blue Nile Basin. *Land Degradation &*

⁶We share a sampling of these papers, clustered by theme, on the project page.

- Development* 31, 15 (2020), 2168–2184.
- [23] Olayiwola Bello, Victor Akinwande, Oluwatoyosi Jolayemi, and Ahmed Ibrahim. 2016. Open data portals in Africa: an analysis of open government data initiatives. *African journal of library, archives & information science* 26, 2 (2016), 97.
- [24] Louise Bezuidenhout. 2017. Technology transfer and true transformation: implications for open data. *Data Science Journal* 16 (2017), 26.
- [25] Louise Bezuidenhout. 2019. To share or not to share: Incentivizing data sharing in life science communities. *Developing World Bioethics* 19, 1 (2019), 18–24.
- [26] Louise Bezuidenhout and Ereck Chakauya. 2018. Hidden concerns of sharing research data by low/middle-income country scientists. *Global Bioethics* 29, 1 (2018), 39–54.
- [27] Louise Bezuidenhout, Sabina Leonelli, Ann Kelly, and Brian Rappert. 2017. Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy* 44, 4 (2017), 464–475.
- [28] Louise Bezuidenhout and Brian Rappert. 2016. What hinders data sharing in African science. In *Presentation and Paper at Fourth Codesria Conference on Electronic Publishing*, Vol. 30. Dakar, Senegal, 1–13.
- [29] Abeba Birhane. 2020. Algorithmic Colonization of Africa. *SCRIPTed* 17 (2020), 389.
- [30] Crystal Biruk. 2018. *Cooking Data*. Duke University Press, Durham, NC.
- [31] Brown Brene. 2010. The power of vulnerability. https://www.ted.com/talks/brene_brown_the_power_of_vulnerability. Accessed: 7 October 2020.
- [32] Winner Chawinga and Sandy Zinn. 2019. Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research* 41, 2 (2019), 109–122.
- [33] Bagele Chilisa. 2019. *Indigenous research methodologies*. Sage Publications, Incorporated, Sage, Los Angeles.
- [34] Elisha Chiware. 2020. Open research data in African academic and research libraries: a literature analysis. *Library Management* 41, 6/7 (2020), 383–399.
- [35] Lorenzo Cini. 2019. Disrupting the neoliberal university in South Africa: The #FeesMustFall movement in 2015. *Current Sociology* 67, 7 (2019), 942–959.
- [36] Danielle Coleman. 2018. 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 (2018), 417.
- [37] Jason Colquitt, Brent Scott, and Jeffery LePine. 2007. Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of applied psychology* 92, 4 (2007), 909.
- [38] Jeff Cornassel et al. 2009. Indigenous storytelling, truth-telling, and community approaches to reconciliation. *ESC: English Studies in Canada* 35, 1 (2009), 137–159.
- [39] Richard Delgado. 1989. Storytelling for oppositionists and others: A plea for narrative. *Michigan Law Review* 87, 8 (1989), 2411–2441.
- [40] Spencer Denny, Blessing Silaigwana, Douglas Wassenaar, Susan Bull, and Michael Parker. 2015. Developing ethical practices for public health research data sharing in South Africa: The views and experiences from a diverse sample of research stakeholders. *Journal of Empirical Research on Human Research Ethics* 10, 3 (2015), 290–301.
- [41] Catherine D'Ignazio and Lauren F Klein. 2020. *Data feminism*. MIT Press.
- [42] Wafeeqa Dinah, Pheladi Lefika, and Bwalya Joseph. 2019. The role of open data in smart cities: exploring status in resource-constrained countries. In *Governance Models for Creating Public Value in Open Data Initiatives*. Springer, Cham, 105–121.
- [43] Nomusa Dlodlo, Thato Foko, Promise Mvelase, and Sizakele Mathaba. 2012. *The state of affairs in internet of things research*. Academic conferences international Ltd, South Africa.
- [44] Lauren Dyll-Myklebust. 2014. Development narratives: the value of multiple voices and ontologies in Kalahari research. *Critical Arts* 28, 3 (2014), 521–538.
- [45] Alebachew Enyew. 2016. Towards Data Protection Law in Ethiopia. In *African Data Privacy Laws*. Springer, Germany, 143–159.
- [46] Open Data for Africa Portal. 2020. Africa Information Highway. <https://dataportal.opendataforafrica.org/>. Accessed: 30 September 2020.
- [47] Daniela Gachago, Janet Condy, Eunice Ivala, and Agnes Chigona. 2014. All stories bring hope because stories bring awareness: students' perceptions of digital storytelling for social justice education. *South African Journal of Education* 34, 4 (2014), 12.
- [48] Tesfay Asgele Gebru, Grmay Kassa Brhane, and Yohannes Gerezhier Gebremedhin. 2021. Contributions of water harvesting technologies intervention in arid and semi-arid regions of Ethiopia, in ensuring households' food security, Tigray in focus. *Journal of Arid Environments* 185 (2021), 104–373.
- [49] Alison Gillwald. 2020. Data, AI & Society. <https://researchictafrica.net/2020/03/10/data-ai-society/>. Accessed: 28 September 2020.
- [50] Kelsey Glennon, Joseph White, Chevonne Reynolds, Ida Risenga, Neville Pillay, Sally Archibald, Kevin Balkwill, Marcus Byrne, Glynis Cron, Hanlie Engelbrecht, et al. 2020. Misinterpretation of why black students do not pursue studies in the biological sciences. *South African Journal of Science* 116, SPE (2020), 1–3.
- [51] Christopher Henshilwood, Francesco d'Errico, Royden Yates, Zenobia Jacobs, Chantal Tribolo, Geoff Duller, Norbert Mercier, Judith Sealy, Helene Valladas, Ian Watts, et al. 2002. Emergence of modern human behavior: Middle stone age engravings from South Africa. *Science* 295, 5558 (2002), 1278–1280.
- [52] Eric Horvitz and Deirdre Mulligan. 2015. Data, privacy, and the greater good. *Science* 349, 6245 (2015), 253–255.
- [53] Carlos Iglesias. 2019. *Open data barometer: African edition report: Open data readiness, use and impact in Africa*. World Wide Web Foundation, Washington DC.
- [54] Hsain Ilahiane. 2017. *Historical dictionary of the Berbers (Imazighen)*. Rowman & Littlefield.
- [55] Samuel Oluoch Imbo. 2002. *Oral traditions as philosophy: Okot p'Bitek's legacy for African philosophy*. Rowman & Littlefield.
- [56] Aaron Inamara and Verena Thomas. 2017. Pacific climate change adaptation: The use of participatory media to promote indigenous knowledge. *Pacific Journalism Review* 23, 1 (2017), 113–132.
- [57] Charlene Janion-Scheepers, John Measey, Brigitte Braschler, Steven Chown, Louise Coetzee, Jonathan Colville, Joanna Dames, Andrew Davies, Sarah Davies, Adrian Davis, et al. 2016. Soil biota in a megadiverse country: Current knowledge and future research directions in South Africa. *Pedobiologia* 59, 3 (2016), 129–174.
- [58] Irene Jao, Francis Kombe, Salim Mwalukore, Susan Bull, Michael Parker, Dorcas Kamuya, Sassy Molyneux, and Vicki Marsh. 2015. Research stakeholders' views on benefits and challenges for public health research data sharing in Kenya: the importance of trust and social relations. *PLoS one* 10, 9 (2015), e0135545.
- [59] Diana Jeater. 2018. Academic Standards or Academic Imperialism? Zimbabwean perceptions of hegemonic power in the global construction of knowledge. *African Studies Review* 61, 2 (2018), 8–27.
- [60] Jane Kaye. 2012. The tension between data sharing and the protection of privacy in genomics research. *Annual Review of Genomics and Human Genetics* 13 (2012), 415–431.
- [61] Moyra Keane, Constance Khupe, and Maren Seehawer. 2017. Decolonising methodology: Who benefits from indigenous knowledge research? *Educational Research for social change* 6, 1 (2017), 12–24.
- [62] Michael Kimani. 2019. 5 reasons why Facebook's new cryptocurrency 'Libra' is bad news for Africa. <https://kioneki.com/2019/06/28/5-reasons-why-facebooks-new-cryptocurrency-libra-is-bad-news-for-africa/>. Accessed: 28 September 2020.
- [63] Ruth Kiraka, Felix Maringe, Watiri Kanyutu, and Emmanuel Mogaji. 2020. University league tables and ranking systems in Africa: emerging prospects, challenges and opportunities. In *Understanding the Higher Education Market in Africa*. Routledge, Oxfordshire, UK.
- [64] Bartha Knoppers, Jennifer Harris, Anne Marie Tassé, Isabelle Budin-Ljosne, Jane Kaye, Mylène Deschênes, and Zawati Ma'n. 2011. Towards a data sharing code of conduct for international genomic research. *Genome Medicine* 3, 7 (2011), 46.
- [65] Patty Kostkova, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, Peter Knight, Corinne Marsolier, Rachel A McKendry, et al. 2016. Who owns the data? Open data for healthcare. *Frontiers in public health* 4 (2016), 1–6.
- [66] Moritz Kraemer, David Pigott, Sarah Hill, Samantha Vanderslott, Robert Reiner, Stephanie Stasse, John Brownstein, Bernardo Gutierrez, Francis Dennig, Simon Hay, et al. 2020. Dynamics of conflict during the Ebola outbreak in the Democratic Republic of the Congo 2018–2019. *BMC medicine* 18 (2020), 1–10.
- [67] John Kurewa. 2016. *African Religion: The quarry of the rock of monotheism*. Upper Room Books.
- [68] René Lemarchand. 2007. Consociationalism and power sharing in Africa: Rwanda, Burundi, and the Democratic Republic of the Congo. *African Affairs* 106, 422 (2007), 1–20.
- [69] Yanni Alexander Loukissas. 2019. *All data are local: Thinking critically in a data-driven society*. MIT Press.
- [70] Gérard Magrin and Geert Van Vliet. 2009. The use of oil revenues in Africa. *Governance of Oil in Africa: Unfinished Business, Paris, IFRI* (2009), 103–163.
- [71] Alex Makulilo. 2016. The context of data privacy in Africa. In *African Data Privacy Laws*. Springer, Cham, 3–23.
- [72] Alicia Martin, Solomon Teferra, Marlo Möller, Eileen Hoal, and Mark Daly. 2018. The critical needs and challenges for genetic architecture studies in Africa. *Current Opinion in Genetics & Development* 53 (2018), 113–120.
- [73] Peter Mataira. 2019. Transforming indigenous research: Collaborative responses to historical research tensions. *International Review of Education* 65, 1 (2019), 143–161.
- [74] Mbazima Simeon Mathebane. 2017. *Towards indigenous social work practice guidelines for assisting African families raising children with Down syndrome*. Ph.D. Dissertation. University of South Africa.
- [75] Rose Mbaye, Redeat Gebeyehu, Stefanie Hossmann, Nicole Mbarga, Estella Bih-Neh, Lucrece Eteki, Ohene-Agyei Thelma, Abiodun Oyerinde, Gift Kiti, Yvonne Mburu, et al. 2019. Who is telling the story? A systematic review of authorship for infectious disease research conducted in Africa, 1980–2016. *BMJ Global Health* 4, 5 (2019), e001855.
- [76] Ian McDougall, Francis Brown, and John Fleagle. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433, 7027 (2005),

- 733–736.
- [77] Daniela Merolla. 2019. Memory and remembrance in Africa and the diaspora. *Afrika Focus* 32, 1 (2019), 5–7.
- [78] Jean Pierre Misago. 2019. Political mobilisation as the trigger of xenophobic violence in post-apartheid South Africa. *International Journal of Conflict and Violence (IJCV)* 13 (2019), 1–11.
- [79] Mucha Mkono. 2019. Neo-colonialism and greed: Africans' views on trophy hunting in social media. *Journal of Sustainable Tourism* 27, 5 (2019), 689–704.
- [80] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology* (2020), 1–26.
- [81] Samuel Maredi Mojapelo. 2020. Storytelling Initiatives and Reading in Schools in South Africa: A Case from Limpopo Province. *Mousaion* 38, 3 (2020).
- [82] Mapule Sheena Moroke. 2015. *Train surfing: the Soweto pastime*. Ph.D. Dissertation. University of the Witwatersrand, Johannesburg.
- [83] Sandra Mudondo. 2020. *The past present dialogue: a recreational park to incorporate memorialisation into contemporary Ugandan culture*. Ph.D. Dissertation. University of the Witwatersrand.
- [84] Nchangwi Munung, Patricia Marshall, Megan Campbell, Katherine Littler, Francis Masiye, Odile Ouwe-Missi-Oukem-Boyer, Janet Seeley, DJ Stein, Paulina Tindana, and Jantina de Vries. 2016. Obtaining informed consent for genomics research in Africa: analysis of H3Africa consent documents. *Journal of Medical Ethics* 42, 2 (2016), 132–137.
- [85] Rosemary Musesengwa, Moses Chimbari, and Samson Mukaratirwa. 2018. A framework for community and stakeholder engagement: experiences from a multicenter study in Southern Africa. *Journal of Empirical Research on Human Research Ethics* 13, 4 (2018), 323–332.
- [86] Leonida Mutuku and Christine Mahihu. 2014. Understanding the impacts of Kenya open data applications and services. (08/2014 2014), 1–58.
- [87] Erisa Sabakaki Mwaka. 2017. *Stakeholder perspectives on informed consent and ethics review of research involving human specimen resource repositories (biobanks) in South Africa: a quantitative study*. Ph.D. Dissertation. University of KwaZulu-Natal.
- [88] Joseph Mwelwa, Geoffrey Boulton, Joseph Wafula, and Cheikh Loucoubar. 2020. Developing open science in Africa: Barriers, solutions and opportunities. *Data Science Journal* 19, 1 (2020), 31.
- [89] Moses Namara, Daricia Wilkinson, Byron M. Lowens, Bart P. Knijnenburg, Rita Orji, and Remy L. Sekou. 2018. Cross-Cultural Perspectives on EHealth Privacy in Africa (*AfriCHI '18*). Association for Computing Machinery, New York, NY, USA, Article 7, 11 pages.
- [90] Willy Ngaka. 2020. The role of communities in Uganda's mother tongue-based education: Perspectives from a literacy learning enhancement project in Arua district. *Applied Linguistics Review* 1, ahead-of-print (2020).
- [91] Lexi Novitske. 2018. The AI invasion is coming to Africa (and it's a good thing). *Stanford Social Innovation Review* (2018).
- [92] Ferdinand N'Sougan Agblemagnon and Simon Pleasance. 1971. The literary and artistic sociology of Black Africa. *Diogenes* 19, 74 (1971), 89–110.
- [93] Wongani Nyangulu, Randy Mungwira, Nginanche Nampota, Oswald Nyirenda, Lufina Tzirizani, Edson Mwinjiwa, and Titus Divala. 2019. Compensation of subjects for participation in biomedical research in resource-limited settings: a discussion of practices in Malawi. *BMC medical ethics* 20, 1 (2019), 82.
- [94] Annalisa Oboe. 2007. The TRC women's hearings as performance and protest in the new South Africa. *Research in African Literatures* 38, 3 (2007), 60–76.
- [95] Academy of Science of South Africa et al. 2019. Mapping the current landscape of Open Science. *Open Data initiatives in Africa: draft report (Online)* (2019).
- [96] Academy Science of South Africa. 2018. The African Open Science Platform: The Future of Science and Science for the Future. (2018).
- [97] Festus Ugboaja Ohaegbulam. 1990. *Towards an understanding of the African experience from historical and contemporary perspectives*. University Press of America, Lanham, Md.
- [98] Emmanuel Okinedo Omovutotu and Tania Mwiza. 2019. *Impacts of colonialism in Africa: A case study of ethnic identity and ethnic conflicts in Burundi*. Bachelor's Thesis. Malmö universitet/Kultur och samhälle.
- [99] Samuel Senayon Olaoluwa. 2008. *Inscribing dispersal: The politics of exile in African poetry*. Ph.D. Dissertation. University of the Witwatersrand, Johannesburg.
- [100] Omwoyo Bosire Onyancha. 2016. Open research data in Sub-Saharan Africa: a bibliometric study using the data citation index. *Publishing Research Quarterly* 32, 3 (2016), 227–246.
- [101] Garry Paterson, Dave Turner, Liesl Wiese, George Van Zijl, Cathy Clarke, and Johan Van Tol. 2015. Spatial soil information in South Africa: Situational analysis, limitations and challenges. *South African Journal of Science* 111, 5–6 (2015), 1–7.
- [102] Okot P'Bitek. 1975. Fr. Tempels' Bantu Philosophy. *Transition* 20, 50 (1975), 66–68.
- [103] Ntshengedzeni Collins Rananga. 2008. *Professionalising storytelling in African languages with special reference to Venda*. Ph.D. Dissertation. University of South Africa.
- [104] Sonja Rasmussen and Richard Goodman. 2018. The CDC field epidemiology manual. <https://www.cdc.gov/eis/field-epi-manual/Publishing-Reference.html>. Accessed: 30 September 2020.
- [105] Michael Schaeppman, Zbyněk Malenovsky, Sander Mücher, Lammert Kooistra, and Wilfried Thullier. 2007. Bridging scaling gaps for the assessment of biodiversity from space. In *The full picture*. Tudor Rose/GEO, Geneva (Switzerland), 258–261.
- [106] Maletle Daniel Sebake. 2013. *Assessing the motivators and barriers of inter-organizational GIS data sharing for address data in South Africa*. Ph.D. Dissertation. University of Pretoria.
- [107] Maletle Daniel Sebake and Serena Coetzee. 2012. Results of three case studies for assessing motivators and barriers of address data sharing in South Africa. *South African Journal of Geomatics* 1, 1 (2012), 32–43.
- [108] Johannes Seroto. 2011. Indigenous education during the pre-colonial period in southern Africa. *Indilinga African Journal of Indigenous Knowledge Systems* 10, 1 (2011), 77–88.
- [109] David Serwadda, Paul Ndebele, Kate Grabowski, Francis Bajunirwe, and Rhoda Wanyenze. 2018. Open data sharing and the Global South—Who benefits? *Science* 359, 6376 (2018), 642–643.
- [110] Edward Shizha. 2014. The indigenous knowledge systems and the curriculum. In *African Indigenous Knowledge and the Disciplines*. Brill Sense, Rotterdam, 113–129.
- [111] Wole Soyinka. 2020. Dionysus in Yorubaland. *Dionysus after Nietzsche: The Birth of Tragedy in Twentieth-Century Literature and Thought* (2020), 161.
- [112] Medadi Ssentanda and Sam Andema. 2019. Reclaiming the space for storytelling in Ugandan primary schools. *Per Linguam: a Journal of Language Learning= Per Linguam: Tydskrif vir Taalaanleer* 35, 3 (2019), 74–91.
- [113] Ciara Staunton, Rachel Adams, Edward Dove, Natalie Harriman, Lyn Horn, Melodie Labuschaigne, Nicola Mulder, Antonel Olckers, Anne Pope, Michèle Ramsay, et al. 2019. Ethical and practical issues to consider in the governance of genomic and human research data and data sharing in South Africa: a meeting report. *AAS Open Research* 2 (2019), 11.
- [114] Louise Swemmer, Rina Grant, Wendy Annecke, and Stefanie Freitag-Ronaldson. 2015. Toward more effective benefit sharing in South African National Parks. *Society & Natural Resources* 28, 1 (2015), 4–20.
- [115] Olufemi Taiwo. 2010. *How colonialism preempted modernity in Africa*. Indiana University Press. 1–352 pages.
- [116] Vanessa Tchamyu and Simplicie Asongu. 2017. Information sharing and financial sector development in Africa. *Journal of African Business* 18, 1 (2017), 24–49.
- [117] Carol Tenopir, Elizabeth D Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS one* 10, 8 (2015), e0134826.
- [118] Abonesh Tesfaye, James Hansen, Maren Radeny, Sebsib Belay, and Dawit Solomon. 2019. Actor roles and networks in agricultural climate services in Ethiopia: a social network analysis. *Climate and Development* 12, 8 (2019), 1–12.
- [119] Bichaye Tessema. 2019. *Sustainable land management in Ethiopia: Response to rural households' tenure insecurity, land degradation and food insecurity*. Ph.D. Dissertation. University of Twente.
- [120] Kudakwashe Tuwe. 2016. The African oral tradition paradigm of storytelling as a methodological framework: Employment experiences for African communities in New Zealand. In *African Studies Association of Australasia and the Pacific (AFSAAP) Proceedings of the 38th AFSAAP Conference: 21st century Tensions and Transformation in Africa*. Elsevier, Australia, 1–13.
- [121] George van Zijl. 2019. Digital soil mapping approaches to address real world problems in southern Africa. *Geoderma* 337 (2019), 1301–1308.
- [122] Susan Veldsman and Ina Smith. 2018. Data driving sustainability—the African open science platform project. *Electronic Publishing* (2018), 9.
- [123] Mark Walport and Paul Brest. 2011. Sharing research data to improve public health. *The Lancet* 377, 9765 (2011), 537–539.
- [124] Aisling Walsh, Ruairi Brugha, and Elaine Byrne. 2016. “The way the country has been carved up by researchers”: ethics and power in north–south public health research. *International journal for equity in health* 15, 1 (2016), 1–11.
- [125] Ibrahim Wuni, Geoffrey Shen, and Robert Osei-Kyei. 2019. Scientometric review of global research trends on green buildings in construction journals from 1992 to 2018. *Energy and Buildings* 190 (2019), 69–85.
- [126] AD Yilma. 2013. Spatially Enabled Africa. In *Proceedings of Global Geospatial Conference*. Publons, Addis Ababa, Ethiopia.
- [127] Nathan Yozwiak, Christian Happi, Donald Grant, John Schieffelin, Robert Garry, Pardis Sabeti, and Kristian Andersen. 2016. Roots, not parachutes: research collaborations combat outbreaks. *Cell* 166, 1 (2016), 5–8.
- [128] Basia Zaba, Clara Calvert, Milly Marston, Raphael Isingo, Jessica Nakiyingi-Miiri, Tom Lutalo, Amelia Crampin, Laura Robertson, Kobus Herbst, Marie-Louise Newell, et al. 2013. Effect of HIV infection on pregnancy-related mortality in sub-Saharan Africa: secondary analyses of pooled community-based data from the network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA). *The Lancet* 381, 9879 (2013), 1763–1771.

