

Open for whose benefit?

Exploring assumptions, power relations and
development paradigms framing the GIZ *Open
Resources Incubator* (ORI) pilot for
open voice data in Rwanda

Daniel Brumund

Table of contents

Abbreviations	i
List of figures	i
Abstract.....	ii
1 Introduction	1
2 Literature review.....	4
2.1 ICTs, big data and development in Rwanda	4
2.1.1 Development and ICTs in Rwanda.....	4
2.1.2 Datafication and the value of big data	6
2.1.3 Big data in development discourses	8
2.2 Open development as an ICT4D approach.....	10
2.2.1 Openness and open development as social praxis	10
2.2.2 Community roles, relations and institutionalisation.....	13
2.2.3 Developmental potentials and paradigms	15
2.3 Relevance to the ComDev research field	17
3 Methodology.....	18
3.1 Analytical framework	18
3.2 Qualitative methods and scope	19
3.3 Reflection on challenges	21
4 Analysis: The GIZ Open Resources Incubator pilot for crowd-sourcing open voice data in Rwanda	22
4.1 Dual context as GIZ idea piloted in Rwanda	22
4.2 Socio-technical process with emergent institutionalisation	26
4.3 Crowdsourcing as global-local open practice	30
4.4 Intermediated usage and indirect developmental benefits.....	33
5 Conclusion.....	35
References.....	39

Abbreviations

AI	Artificial intelligence
CA	Capabilities approach
CC	Creative Commons
ComDev	Communication for development
CV	Common Voice (Project by Mozilla)
DSSD	Digital Solutions for Sustainable Development (GIZ programme in Rwanda)
DU	Digital Umuganda (start-up in Rwanda)
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
ICT	Information and communication technology
ICT4D	Information and communication technology for development
IP	Intellectual property
ML	Machine learning
NCPWD	National Council for Persons with Disabilities
NDP	National development plan
NIRDA	National Industrial Research and Development Agency
OD	Open development
OECD	Organisation for Economic Cooperation and Development
ORI	Open Resources Incubator
RDB	Rwanda Development Board
RISA	Rwanda Information Society Authority
RPF	Rwandan Patriotic Front
RURA	Rwanda Utilities Regulatory Authority
SDG	Sustainable Development Goal
UN	United Nations
WEF	World Economic Forum

List of figures

Figure 1: Four elements of an open process adapted from Smith and Seward (2017) ..	19
Figure 2: Participants of the GIZ/Mozilla hackathon at kLab in Kigali	24
Figure 3: Winning team (middle) and founders of start-up <i>Digital Umuganda</i>	25
Figure 4: Mozilla's <i>Common Voice</i> platform for voice recordings in Kinyarwanda	26
Figure 5: Mozilla's <i>Common Voice</i> platform for voice recordings in English	27
Figure 6: <i>Common Voice</i> 'Sentence collector' for adding and reviewing text data	27
Figure 7: Data collection events at universities organised by <i>Digital Umuganda</i>	31

Abstract

Since February 2019, the Kigali-based start-up *Digital Umuganda* has been coordinating the crowdsourcing of the first openly available voice dataset for Rwanda's official language Kinyarwanda. This process originated from a pilot project of the *Open Resources Incubator* (ORI), an emergent service designed by GIZ staff and the author as consultant. ORI aims to facilitate the collective provision of open content, thereby affording previously inaccessible opportunities for local innovation and value creation. It promotes the community-based stewardship of open resources by (inter-)national actors who share responsibilities for their production, distribution and use. ORI's pilot project cooperates with Mozilla's team behind *Common Voice*, a platform to crowdsource open voice data, and has attracted Rwandan public and private actors' interest in voice technology to improve their products and services.

Informed by research on ICTs, datafication and big data in development discourses, and using the ICT4D approach 'open development' as its analytical lens, this thesis examines inherent conceptual aspects and socio-technical dynamics of the ORI pilot project. An in-depth analysis of qualitative data gathered through participant observation, interviews and focus groups explores assumed developmental benefits which international and Rwandan actors involved in the project associate with open voice data, power relations manifesting between these actors as well as underlying development paradigms.

The analysis shows how the project established a global-local datafication infrastructure sourcing voice data from Rwandan volunteers via technically, legally and socially formalised mechanisms. By placing the dataset in the public domain, the decision as to how it will be used is left to the discretion of intermediaries such as data scientists, IT developers and funders. This arrangement calls into question the basic assumption that the open Kinyarwanda dataset will yield social impact because its open access is insufficient to direct its usage towards socially beneficial, rather than solely profit-oriented, purposes. In view of this, the thesis proposes the joint negotiation of a 'stewardship agreement' to define how value created from the open voice data will benefit its community and Rwanda at large.

Keywords: Big data, Open development, Datafication, Crowdsourcing, Rwanda

1 Introduction

“If we begin to view data as the raw material for the information age, the question of data ownership, access and utilization becomes one of prime importance; policies need to be in place to democratise ownership and use of data. If we don't address these issues now, the future will bring a much bigger divide with far reaching consequences than we experienced in the past, when it was mostly about connectivity and access.”

Patrick Nyirishema (2017)
Director-General, Rwanda Utilities Regulatory Authority (RURA)

“Raw data’ is both an oxymoron and a bad idea.”
Geoffrey C. Bowker (cited in Gitelman, 2013)

It is mid-December 2019 at the African Leadership University in the Rwandan capital Kigali. Some thirty students sit in front of laptops reading out and recording sentences in Kinyarwanda using Mozilla's crowdsourcing platform *Common Voice*. They participate in a data collection event by *Digital Umuganda* (DU), a Kigali-based start-up working towards creating an open Kinyarwanda dataset. Supported by Mozilla, DU plans to develop the first openly available Kinyarwanda speech model. This sparked interest among public and private actors in Rwanda, such as smartphone manufacturer *MaraPhone* who wants to offer devices with Kinyarwanda voice-recognition, or public officials who wish to improve accessibility of the eGovernment portal Irembo via voice-interaction. From promoting entrepreneurship to simplifying access to services, these actors associate a wide range of benefits with open voice-recognition technology in Rwanda.

The founding of DU and its efforts to build an open Kinyarwanda dataset have been inspired and supported through a collaboration between Mozilla and the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), particularly a new service called the 'Open Resources Incubator' (ORI). ORI emerged from a GIZ-internal innovation programme as part of which it has been designed since early 2019 by a team of GIZ staff and me as consultant. The aim of ORI is to facilitate open access to resources such as voice data, thereby contributing to development by affording previously inaccessible opportunities for local innovation and value creation. At its centre is a community-building approach focused on identifying, connecting and supporting actors who share responsibilities for the stewardship of the open resource.

Supporting the community-based production and management of an open Kinyarwanda dataset is ORI's pilot project to test its approach. By February 2020, this had resulted in a growing network of actors including DU who coordinates the voice data collection, Mozilla who provides the technical platform and ML expertise, a Kigali-based GIZ programme which supports DU financially, Rwandan universities whose students

volunteer as contributors of voice data, as well as public institutions and private companies interested in using voice technology.

While the crowdsourcing is on-going, certain assumptions already exist among these actors concerning developmental benefits of open voice data, and varying relations manifest between them vis-à-vis technological, financial or social dependencies. My thesis aims to investigate these assumptions and power relations framing the ORI pilot in Rwanda, as well as any underlying development paradigms. To explain why I consider this a worthwhile research topic, I need to introduce the broader context of big data, ICT and development at the intersection of which the ORI pilot operates.

A core idea behind the ORI pilot in Rwanda is that opening access to Kinyarwanda voice data may help lower barriers for public and private actors to utilise and benefit from an increasingly valuable technology – voice interaction; and that this, in turn, helps promote local innovation and value creation. Why would this matter for Rwanda? Technological advances in voice recognition provide opportunities to simplify how people engage with technology or access information in sectors including health care, education or governance, but these opportunities are unequally distributed. While market research indicates the rise of a multi-billion industry around voice interaction (Grand View Research, 2018), the global South, particularly Sub-Saharan Africa, is barred from even exploring potentials of voice technology because African languages are largely underrepresented digitally and respective voice datasets do not yet exist (Dahir, 2018).

Efforts by tech corporations such as Google to digitise African and Indian languages for reaching the ‘next billion’ Internet users (Bellman, 2017) must be viewed critically. Far from altruistic, these efforts are rather driven by the market potential seen in expanding fee-based voice-interaction services. This makes the technology accessible only to those who can for it pay while privatising rewards gained from the ‘datafication’ (Van Dijck, 2014) of essential aspects of human interaction: people’s voices and languages.

The datafication of human behaviour and corporate appropriation of these data for private value extraction exemplifies a process which has been criticised as ‘data extractivism’ (Morozov, 2018b) and ‘data colonialism’ (Couldry & Mejias, 2019b, 2019c). The exploitative character of this process is concealed by dominant discourses framing data as the ‘the new oil’ (Economist, 2017b) of the digital economy. Yet, as IT professor Geoffrey C. Bowker stresses in his introductory quote, the notion of data as being ‘raw’ is an oxymoron: Unlike natural resources, data are always socially constructed in specific contexts and for specific purposes.

The digitisation of Sub-Saharan languages by corporations in the global North for their profitmaking leads to a monopolisation of power by privatising ownership and usage of these (big) data. Countries in the global South are particularly vulnerable to the negative effects of this process which creates new dependencies and forecloses them from exploring the potentials of big data or translating them into local benefits (Graham, 2019; Ojanperä et al., 2019). It mirrors the notion expressed in the introductory quote by Patrick Nyirishema, director-general of Rwanda's national regulatory agency RURA, that democratic ownership, access and utilisation of data are increasingly important issues for the global South – and that neglecting them risks deepening digital divides.

The ORI pilot in Rwanda is an attempt to counteract such risks and processes through producing and opening access to a Kinyarwanda dataset and speech model. The purpose of my thesis is to explore this process along the following research questions:

- **Main question:** What are the main assumptions by actors involved in the ORI pilot regarding the developmental benefits of open voice data, and how do these influence its implementation in Rwanda?
- **Sub-question (1):** What power relations between the involved actors – in terms of motivations, roles and dependencies – manifest throughout the ORI pilot?
- **Sub-question (2):** What dominant development paradigms are underlying the ORI pilot in Rwanda?

To investigate these questions, I elaborate a case study of the ORI pilot in which I analyse qualitative data gathered via three research methods: firstly, participant-observation in the ORI pilot by myself as researcher, consultant and part of the project team; secondly, semi-structured interviews with main actors involved in the pilot; and, thirdly, focus group discussions of main findings with DU and the GIZ staff behind the ORI project.

As theoretical lens for the analysis, I use 'open development' (OD) which is an ICT4D approach concerned with how ICT-facilitated open production, distribution and usage of information may affect development processes (Reilly & Smith, 2013). Examples include initiatives promoting open data, open government or open educational resources as means to advance development in different contexts. They align with the OD premise that openly networked ICTs afford people opportunities to collaboratively produce, share and use information in ways that may help them improve their lives (Reilly & Alperin, 2016).

An important argument of OD research is that it is less tech-centric, as in what content should be open, and more people-centric, as in how openness unfolds in practice (Reilly & Smith, 2013). Underlying OD is an understanding of openness as 'social praxis'

involving three ICT-facilitated processes: open production, open distribution and open consumption of content (Smith & Seward, 2017). Some scholars argue these processes help decentralise knowledge-heavy processes, redistribute power over information and spur innovation (Reilly & Smith, 2013; Smith et al., 2011). Others caution that open processes replicate existing power imbalances and inequalities within society which they themselves cannot affect (Singh & Gurusurthy, 2013).

Against this background, OD with its inherent debates about the potentials and limitations of ICT-enabled openness offers a valuable lens for critically investigating assumptions, power relations and development paradigms framing the ORI pilot in Rwanda – at the heart of which lies the open process of crowdsourcing a Kinyarwanda voice dataset.

2 Literature review

“Knowledge, innovation and the capacity to communicate are core elements of human development and contribute to its improvement. They are also the building blocks and core outputs of the networked environment. [...] It will not be the solution for (or cause of) all aspects of human development. But the new set of effective human action that it makes feasible fundamentally reshapes the problems, solutions, and institutional frameworks of human development.”

Yochai Benkler (2013, p. ix)

Two broad areas of academic research are relevant for informing my thesis: firstly, ICTs in the context of development in Rwanda, and how big data and underlying datafication processes relate to it; and secondly, open development as an ICT4D approach, including its social aspects and dominant development paradigms. This chapter outlines current debates in each area. Benkler’s introductory quote is a reminder that while ICT-networked environments offer opportunities to promote knowledge exchange and innovation, they are neither solution nor cause of development but, at best, useful means for supporting it.

2.1 ICTs, big data and development in Rwanda

2.1.1 Development and ICTs in Rwanda

Within twenty-five years, Rwanda evolved from a country ravaged by civil war and genocide in the mid-1990s towards what the World Bank (2019) calls a remarkable social, political, and economic renaissance. Crisafulli and Redmond (2012) view Rwanda as an example for poverty reduction through “private sector-led development, decentralised government, transparency and accountability at all levels” (p. 5). The country’s transition has been credited to the post-genocide government led by the Rwandan Patriotic Front (RPF) under President Paul Kagame, whom Crisafulli and Redmond liken to a ‘CEO of Rwanda, Inc’ boldly executing his vision of a free market, technology-based Rwandan economy.

Walking through central Kigali, one witnesses Rwanda's impressive turn-around. Yet, the above account is one-sided in its failure to question the regime's authoritarianism or the inequality entrenched by its economic model. In a less adulatory account of Rwanda's post-war development, Thomson (2018) argues that the country's authoritarian leadership "limits political rights and freedoms in exchange for national security via a mix of overt repression and subtle co-optation" (p. 3). Longman (2011) adds that Kagame pursues a strategy of 'performance legitimization' assuming that prosperity will justify his heavy-handed means. Yet, this prosperity hardly spreads beyond Kigali. Moreover, top-down mandated measures such as requiring citizens to wear shoes in public or obliging them to participate in monthly *Umuganda* public labour programmes may be sensible from a safety or communal perspective, but they instil a socio-political climate of submission which is socially monitored for compliance (Purdeková, 2016). While the government defends its 'benign authoritarian rule' as the basis of national unity and economic development, Longman (2011) argues it harms Rwanda's long-term stability by preventing "the public from expressing its interests through productive, peaceful political means and [...] the regime from benefiting from the contributions of much of the population" (p. 27).

Furthermore, the government's promotion of pro-private economic growth fails to reduce inequality in Rwanda. Instead it mostly benefits those close to the RPF-led regime whose private sector involvement resembles 'developmental patrimonialism' (Booth & Golooba-Mutebi, 2012). This is problematic for two reasons: firstly, wealth remains concentrated in cities, particularly Kigali, while most people still reside in rural areas, with 40% living in poverty and 16% in extreme poverty (BTI, 2018). And secondly, the economic model built on the premise of RPF rule hinges on the stability of political conditions (Takeuchi, 2019). For now, Kagame secured a third term as president by changing the constitution. Recent political assassinations leave uncertainty as to what his authoritarian leadership may bode for Rwanda's future (Sebarenzi & Twagiramungu, 2019).

An important pillar of Rwanda's private sector-led development is its focus on ICTs. Since adopting the first national development plan (NDP) in 2000, the government has been determined to transform the economy from subsisting on agriculture into a knowledge-based economy (Friederici et al., 2013; Nibeza, 2015). Investments in ICTs are seen as an opportunity to overcome Rwanda's disadvantages of being a landlocked, resource-poor country. To this extent, the latest NDP – the *Smart Rwanda 2020 Master Plan* (MYICT, 2015) – outlines the aim to "secure and accumulate knowledge competency, as the driver of productivity and economic growth" (p. 14).

As a cross-governmental priority, this process has largely been driven by President Kagame himself who has been dubbed a 'digital president' (Basaninyenzi, 2013) drawing

Rwandan ICT professionals from the diaspora into government with the promise of shaping the country's future (Gagliardone & Golooba-Mutebi, 2016). Moreover, the regime extended its patrimonialism to negotiations with ICT companies, encouraging them to invest in Rwanda through public-private partnerships (Booth & Golooba-Mutebi, 2012). The government also embraced a wide range of externally funded initiatives to help develop its ICT infrastructure, digitise its service delivery or provide ICT-focused capacity-building measures – e.g. by attracting Carnegie Mellon University to Kigali, or setting up innovation hubs such as the Japanese-funded Knowledge Lab (kLab) and the German-funded Digital Transformation Centre (Gagliardone & Golooba-Mutebi, 2016; Kaliisa, 2019; Ntale et al., 2013).

The government's progressive framing of ICTs and its investments in ICT-focused policies and projects has become, as Gagliardone and Golooba-Mutebi (2016) argue, a "source of legitimacy [of public officials] within a context where political competition is limited" (p. 5). The top leadership has delegated power to public officials for implementing ICT policies independently and without the need for further political approval. Concerns remain, though, regarding the use of ICTs for citizen surveillance, the blocking of oppositional internet content or military influence over ICT regulations (Freedom House, 2018). It shows that the regime's emphasis on "the Internet as a tool for economic development also meant de-emphasising its potential as a tool for political change" (Gagliardone & Golooba-Mutebi, 2016, p. 17).

At the same time, a look at three ICT-related indicators reveals the limited impact of ICTs on Rwandan society at large: firstly, ICT accessibility remains low with internet penetration of 21.8% (ITU, 2018) and many Rwandans speaking only Kinyarwanda, thus rendering most online content unintelligible (Samuelson & Freedman, 2010); secondly, for half the population internet-ready mobile devices remain unaffordable (NISR, 2018); and thirdly, skills to meaningfully use ICTs remain limited as indicated by digital literacy rates of 8.9%, with a stark divide between urban (26%) and rural (4.6%) areas (NISR, 2018).

Under pressure to transform its economy (Mann & Berry, 2016), the Rwandan regime has been legitimising its authoritarian practices with a discourse of progress, also through ICTs. This explains the heavy investment in expanding ICT infrastructure and skills by brokering public-private partnerships and accepting dependence on external actors. Yet, concerns remain regarding the societal benefit of ICTs and their use for political purposes.

2.1.2 Datafication and the value of big data

In the introductory quote (on pg. 1), director-general of the Rwandan regulatory authority RURA, Patrick Nyirishema (2017), expressed a concern that unequal access, ownership

and usage of data may foster new digital divides. To understand this concern – shared by the team behind ORI and its Rwandan pilot – it is necessary to explore the processes and discourses framing the production and value of data today.

Referred to by *The Economist* as the “fuel” (2017a) of the digital era and the “world’s most valuable resource” (2017b), data are arguably the core driver of today’s digital economies. Their value can be best understood from a technological and informational perspective. Advances in data storage, computational power and means of transmission enabled an increase in data volume (of information), variety (of sources) and velocity (of creation, storage and dissemination) – the ‘three Vs’ of big data (Hilbert, 2016). This technological progress amplified a process Mayer-Schönberger and Cukier (2013) coined as ‘datafication’: It describes the conversion of (social) interactions into quantifiable data that can be collected, analysed, transferred and monetised. The informational value of these big data lies in providing “key inputs to the various modes of analysis [...] in order to understand and explain the world we live in, which in turn are used to create innovations, products, policies and knowledge that shape how people live their lives” (2014, p. 1). Datafication, thus, combines two processes which Meijas and Couldry (2019) summarise as “the transformation of human life into data through processes of quantification, and the generation of different kinds of value from data.” Its outcomes – positive or negative – have less to do with data per se but with “how and by whom [data are] systematically collected and used” (n.p.).

The use of data is influenced by two dominant discourses concerning its origin and value. Firstly, the discourse framing data as ‘oil’ or ‘raw’ positions it as a natural, objective resource. Van Dijck (2014) cautions against viewing data as objective representations of human experience because they are always collected and processed in specific contexts and for specific purposes. Data are socially constructed which is why Gitelman (2013) describes the notion of ‘raw data’ as an oxymoron. Scholz (2018) argues further that the “data-as-oil analogy side-steps evaluation of any misappropriation or exploitation that might arise from data use and processing by adopting an analogy that presumes the history of the data prior to collection is irrelevant.”

Secondly, the discourse framing data as valuable information simply to be extracted from human interactions without much concern for its ‘producers’ mirrors a flawed narrative of economic value. Economist Mazzucato’s (2018a) argument that today’s economies regard the extraction of value more highly than its creation extends to the digital world: e.g. in the way Google or Facebook extract value from data generated by users of their services for immense corporate profit – raising the question to what extent they may socialise data production while privatising its rewards (Mazzucato, 2018b).

These discourses manifest in power relations between those who produce the data and those who collect, own and profit from them. In this regard, Morozov (2018b, 2019) indicates a conflict between 'data extractivism' driven by corporations' dependence on new data sources, and 'data distributism' promoting data as collectively produced and socially useful resources rather than privatised commodities. The large-scale extraction of data and corporate profit-making from them results in asymmetric power relations between data producers and data users. These easily translate into knowledge asymmetries and dependencies in terms of who is able to utilise, innovate with and benefit from these data.

Thatcher et al (2016) call it a form of 'capitalist expropriation and dispossession' which gives rise to 'data colonialism'. Couldry and Mejias (2019b) describe its underlying double logic as "data abstract[ing] life by converting it into information that can be stored and processed by computers; and appropriat[ing] life by converting it into value for a third party" (p. xiii). The problem with this logic is not the collection of data as such but that the process is externally driven "on terms that are partly or wholly beyond the control of the person to whom the data relates." (p. 5). This paves the way for exploitation.

How to counter such exploitation has been at the centre of growing public and academic debates. One difficulty lies in balancing individual and collective rights over data. Some scholars call for reimbursing citizens for their data (e.g. Arrieta-Ibarra et al., 2018; Lanier, 2014). Others argue that attempts to solve the issue at individual level are insufficient because they condone the underlying practices of data appropriation and usage outside of citizens' control (Couldry & Mejias, 2019a; Zuboff, 2019).

Another way is suggested by Mazzucato's (2018a) who calls for recognising economic value – also of data – as created collectively between public, private and civil society actors. She suggests data could be stored in public repositories with strong conditions attached to their use, particularly by corporations. Similarly, Morozov (2018a, 2019) argues for 'social rights' to data as he calls on public institutions to manage data pools and support their use by projects promising social impact. This could promote innovation by enabling a wide range of (previously excluded) actors to utilise data in socially beneficial ways.

2.1.3 Big data in development discourses

How do big data factor into development discourses, particularly pertaining to the global South? Several reports by international agencies emphasise the potential of big data while stressing mainly technical challenges. For instance, an early UN Global Pulse (2012) publication views big data as valuable informational inputs for decision-making but recognises technical obstacles such as lacking data access or quality. A more recent World Bank (2018) report on 'data-driven development' argues big data may improve

service delivery, listing data protection as a main challenge. What these reports lack, however, is a thorough interrogation of structural challenges posed by big data and the implications for the global South.

A growing body of critical research offers insights in this regard. For instance, Arora (2016) finds biases in how big data are framed as an 'instrument of empowerment' for marginalised citizens in the global South. She argues that idyllic notions of big data, e.g. supporting more efficient public service delivery for the poor, neglect how such efficiencies often come at the cost of these citizens' privacy or their lack of control over underlying data infrastructures. Hence, debates about big data in the global South should be guided by the aspirations, values and practices of citizens. Yet, most discourses about big data and development, as Unwin (2017) highlights, focus on how they "can be used *for* the poor, rather than *by* the poor." For marginalised people to be able to benefit from big data, they need to be capable to use them to "influence their own lives, and yet almost by definition poor people have neither the wherewithal to afford the technology to use big data effectively, nor the expertise to analyse [them]" (p. 166).

Another critique is raised by Couldry and Yu (2018) who show how reports by the WEF (2012), UN (2012) and OECD (2015) frame big data as raw 'data exhaust' from people's lives that have no value or ownership status unless they are collected and repurposed by public or private actors. They criticise this framing for helping to protect datafication processes "from ethical questioning while endorsing the use and free flow of data within corporate control." It promotes the view that "large-scale data collection facilitates data-driven social transformation without reference to people at all" (p.5).

A lack of capacities to utilise big data or to control underlying datafication infrastructures gives rise to 'big data divides' (Andrejevic, 2014; boyd & Crawford, 2012) between those with access, ownership and skills concerning big data – and those without. Countries in Sub-Saharan Africa may become 'information societies' by making data available while falling short of using these data as knowledge for their own benefit. This resembles a process Carmody (2013) called 'thintegration' in reference to another 'ICT revolution': the spread of mobile phones in Africa. He argued that mobile connectivity facilitates only a 'thin' form of integration into the global economy as long as mobile devices and the services offered through them are largely imported. The irony is that African countries are main resource suppliers to produce mobile phones but benefit only marginally from these devices while most profitable mobile services are owned by corporations in the global North.

In the context of big data, there is a similar risk that African countries are 'thintegrated' into digital economies if they become data suppliers while remaining dependent on firms

elsewhere to access or utilise the data sourced from them. This would flip ‘big data for development’ into ‘development for big data’ in the interest of market expansion – reflecting a critique levelled against ICT4D by Unwin (2017). Moreover, it would exacerbate inequalities between global North and South by expanding ‘informational capitalism’ (Cohen, 2013; Pieterse, 2010) through big data and turning development interventions into mere “by-products of [its] large-scale processes” (Taylor & Broeders, 2015, p. 229).

Taylor and Broeders (2015) identify two trends in this regard: firstly, the emergence of public-private partnerships for datafication in the global South which turn companies into powerful development actors; and secondly, a shift from state-collected data towards commercially generated data through citizen surveillance. Both trends result in a lack of accountability for data collection by private actors and a lack of citizens’ control over data about them. Hence, it is important to interrogate how big data in development initiatives introduce new power relations – particularly when government and corporate interests intersect, as in Rwanda. The regime’s reliance on public-private partnerships also extends to big data which the *National Data Policy* describes as the basis of an “innovation-data-enabled industry to harness rapid social economic development” (MYICT, 2017).

This illustrates why Milan and Treré (2019) are right to reject universalist views of big data as an objective resource detached from social, political or economic contexts. Big data are highly contextual, socially constructed and may well reinforce inequalities. Hence, Couldry and Mejiias (2019b) call for re-envisioning data as a “resource whose value can be sustained only if locally negotiated, managed, and controlled” (p. 196). They stress the need to find collective, legally protected means for organising the “social use of data [...] for collective benefit and not corporate privilege” (p. 207). One such means may be by opening the production, distribution and usage of data which is at the heart of open development explored in the next chapter.

2.2 Open development as an ICT4D approach

2.2.1 Openness and open development as social praxis

From open source software to open government or open data, the term ‘open’ has been applied to a wide range of domains. It builds on the notion that creating valuable knowledge-based products or services does not need the enclosure of information but that it can be promoted by openly producing, sharing and utilising information (Mizukami & Lemos, 2008; Pomerantz & Peek, 2016). ICT has been a crucial facilitator of such open practices, particularly the internet with its opportunities for peer-communication and cooperation (Benkler, 2006, 2017c). Consequently, ICT-facilitated openness and collaboration have

arguably had profound, even transformative, effects on social, political and economic processes (Smith & Seward, 2017).

Open development (OD) picks up on these effects: Introduced as an ICT4D approach by the International Development Research Centre in 2011, it describes a field of action and research concerned with whether and how ICT-enabled open practices help advance development processes (Reilly & Smith, 2013; Smith, 2015). OD proponents assume that ICT-facilitated networks afford people opportunities to openly mobilise and organise resources in ways that are less hierarchical and disintermediate knowledge-heavy processes such as governance, education or innovation (Reilly & Alperin, 2016).

There is ambiguity, however, as to what ‘open’ means in different contexts (Pomerantz & Peek, 2016). According to Smith and Seward (2017), the term is commonly used to refer to open artefacts (e.g. open data) or open social institutions (e.g. open government). These reference points for openness are problematic because their clear-cut definitions (open vs. closed) are misleading: “The boundaries of openness in *practice* are much blurrier than the theoretical boundaries of any one definition of an open artefact or social institutions” (n.p.). Therefore, they argue for shifting the understanding of openness in OD from artefact (what is open) to practice (how openness unfolds). Openness as ‘social praxis’ is expressed through a combination of three processes: open production, open distribution and open consumption of content.

Open production refers to the “distributed collective intelligence of people to accomplish something, such as co-creating a digital artefact, solving a problem, or completing a task” (Smith & Seward, 2017). There are two main open production practices: firstly, peer production which involves the open creation and sharing of information and knowledge through people whose decentralised cooperation is facilitated by ICT-based networks (Benkler, 2006, 2017c). Often organised as commons, it is self-governed, driven by non-monetary interests and resulting in non-proprietary outputs (Benkler, 2017a; Benkler et al., 2015). And secondly, crowdsourcing which differs from peer production in that it is centrally organised, e.g. by an institution, to outsource certain tasks to a group of people. Its goals are usually pre-defined by external actors (Benkler, 2017c).

Open distribution means sharing and republishing content such as data for use by others. This content is non-proprietary (i.e. shared for free), non-discriminatory (i.e. free to access and use) and typically offered via digital platforms. It requires legal frameworks that protect its openness, such as *copyleft* licences including GNU General Public Licences or Creative Commons (CC). These subvert intellectual property (IP) law by adapting it to offer flexible degrees of openness (Linksvayer, 2013). Open distribution

requires a shift from closed IP systems which privatise content value towards open IP systems which promote the value of cooperation and sharing. In their exploration of IP systems in Nigeria, De Beer and Oguamanam (2013) found that such a shift needs new formats of IP training as strong IP protection is often seen as necessary for development.

Lastly, open consumption describes the set of practices for using open content. These practices differ according to the type of content but generally include retaining, re-using, revising and/or remixing it (Smith & Seward, 2017). However, 'consumption' is a rather passive process. Therefore, it might be more fitting to refer to open 'usage' which frames participants as active in producing, distributing and utilising content (Benkler, 2000).

This process view, as Smith and Seward (2017) argue, allows for a more bottom-up, practice-oriented understanding of openness rather than top-down definitions of which content ought to be open. Bentley (2017) adds that it also helps explore how and why particular arrangements of people, technology and content might offer opportunities for development. In OD practice, open processes take the form of 'socio-technical systems' (Bentley et al., 2019; Smith & Reilly, 2013) which involve both social structures (i.e. people collaborating in the open production, distribution and/or usage of content) and technical structures (i.e. ICTs facilitating these open processes).

Some OD scholars argue that people are free to participate in open processes because of their non-discriminatory and cost-free character (Reilly & Smith, 2013). Others stress how open processes discriminate against materially deprived communities which lack sufficient means to participate in them due to social, economic or political constraints (Buskens, 2013; Elder et al., 2013). For instance, in a study on *Map Kibera*, a project engaging citizens in Nairobi's largest informal settlement to peer-produce a digital map of their community, Berdou (2017) finds that the usually non-paid character of open production practices ignores participants' concerns to meet daily needs, find training or employment. OD initiatives should, therefore, invest in people's long-term involvement rather than one-off initiatives. Berdou also reveals tensions between the project's technical goal of quickly peer-producing a digital map and its social goal of community engagement which is time-consuming. She recommends analysing "power relations for understanding how technology, information co-creation, and community dynamics intersect" (29). If such issues are disregarded, openly produced and distributed content may rather benefit those already skilled to use it, thus empowering the already empowered (Gurstein, 2011) and increasing relative inequality (Heeks & Shekhar, 2019).

Through a systematic review of OD initiatives, Bentley and Chib (2016) detect a growing interest in development impacts of openness but also a "severe neglect of poor and

marginalised perspectives” (p. 12). Investigating this further, Bentley et al (2019) find that technocentric, normative ideals of openness held by OD practitioners and researchers foreground hypothetical outcomes and ignore local realities. They argue that openness as social praxis has limited impact on structural transformation and power redistribution. While technology is an important facilitator of open processes, overly technocentric views neglect how “social structures and competing interests affect the context and use of open environments” (803). They recommend a critical lens for analysing OD initiative to explore “power-knowledge interplays in open socio-technical processes” (p. 788) and to consider “the empowering and disempowering effects of openness, as well as the [intentions] of actors in shaping open [...] theory and practice” (p. 803).

2.2.2 Community roles, relations and institutionalisation

As socio-technical systems, open processes engage people who use ICTs to openly produce, distribute or use information resources. These people form a community around the open process in which they perform certain roles and, thereby, form varying relations with each other. Exploring research concerning these aspects will help explore roles and relations between actors involved in the ORI pilot.

In terms of community roles in open processes, a distinction can be made between roles within the community and roles of communities as a whole. Regarding the former, a useful typology comes from the open source domain (Nickolls, 2017) distinguishing four types of actors necessary for sustainable communities: maintainers who manage the project and take responsibility for its quality and governance; contributors who offer time and experience to contribute to a project; consumers or users who utilise outputs for their goals; and sustainers who are concerned with the project’s future viability. This four-role typology offers practice-oriented categories for mapping which roles need to be fulfilled for forming viable communities around open processes.

Actors in these communities often act as intermediaries helping users with the uptake of open content (Reilly & Alperin, 2016). Schalkwyk et al (2015, 2016) illustrate how chains of different intermediaries exist who promote open content (e.g. advocacy organisations), who aggregate and distribute it (e.g. tech organisations), or who adapt it for local usability (e.g. CSOs). In OD initiative, Reilly and Alperin (2016) identify five models of such intermediation: ‘decentralised’ focusing on open access to content; ‘arterial’ focusing on information flows between content providers and users; ‘ecosystem’ focusing on institutional relationships to produce value from content; ‘bridging’ focusing on mediators adapting content into usable formats; and ‘communities of practice’ focusing on commons-based management of content.

This indicates how the assumption that ICT-facilitated networks disintermediate knowledge-heavy processes is too simplistic. Whenever open content first needs to be made useable, intermediation is required. This results in new power relations e.g. visible in emergent business models built on intermediating open data (Janssen & Zuiderwijk, 2014). Exploring telecentres as infomediaries in India, Singh and Gurumurthy (2013) found that they built businesses on commodifying and monetising information, thereby increasing the local communities' dependency and eroding their autonomy.

Regarding communities as a whole, Reilly and Alperin (2016) point to stewardship as an important community role in open processes. It describes the accountable management and safeguarding of goods, often on behalf of others (Plotkin, 2014). Reilly and Alperin (2016) argue that communities forming around open processes become stewards not just of the open content but of entire process of producing, distributing and using it. How this stewardship is arranged – in terms of motivations, agreements or mechanisms – shapes the “distribution of responsibilities for production, maintenance, and the use of [open resources], as well as flows of [...] social value that emerge from them” (p. 54).

Concerning the relations between the actors in these communities, Bentley (2017) presents ‘accountability’ as a reflective framework for exploring whether people’s interests, aspirations and capacities are respected in OD initiatives. While generally used for holding institutional actors to account, Bentley argues that ‘accountability’ can be applied to OD more broadly by investigating open processes along three purposes: ‘participation’ to explore the options people have to engage in open processes; ‘responsiveness’ to explore in how far open processes respond to people’s needs and capacities; and ‘obligation’ to explore rules and regulations, as well as actors’ roles and responsibilities in open processes.

This leaves the question how communities around open processes can be sustained. In addition to ensuring key roles are fulfilled and accountable relationships exist, research by Singh and Gurumurthy (2013) points to the need for institutional anchoring. According to them, ICT-enabled networks offer opportunities for institutionalising open processes that involve actors at different socio-political levels “to share competencies, resources, and outcomes” (p. 189). This could take the form of a ‘network public’ operating at three socio-political levels: at ‘micro-local level’ to support communities between municipalities and civil society with connectivity, capacity-building and basic digital tools; at ‘meso-social level’ to facilitate communities of public, private and civil society actors that become responsible for providing various open resources for public use and benefit; and at ‘macro-institutional level’ to establish policy and regulatory structures which support developmental efforts at micro- and meso-levels.

2.2.3 Developmental potentials and paradigms

OD is a phenomenon of the 'network society' (Castells, 2010) and the widespread diffusion of ICTs. Conceptually, it builds on Benkler's (2006) notion of the 'networked information economy' as a system for peer-producing information by distributed individuals through non-market oriented, ICT-facilitated means. A core assumption of OD is that ICTs alter and decentralise socio-economic power relations concerning information and knowledge by way of opening their production, distribution and usage. Beyond providing access to ICTs, this offers opportunities for development by addressing "digital inequality among access holders to make meaningful use of that access" (Reilly & Smith, 2013, p. 21).

It shows the proximity of OD to the capabilities approach (Nussbaum, 2013; Sen, 2001), which views development as the nurturing of people's capabilities and the expansion of their freedoms to live self-determined lives – considered as both means and end of development. Reilly and Smith (2013) argue that OD can also both constitute development by affording people opportunities to participate in open processes, and advance development by expanding peoples' capabilities with increased access to digital resources and networking opportunities. However, Bentley and Chib (2016) caution against viewing OD as development in itself because openness does not guarantee progressive outcomes. Its impact may also be regressive e.g. when open practices such as crowdsourcing benefit the organisers more than the contributors, thereby entrenching exploitation and the concentration of economic power (Kleemann et al., 2008).

An important capability associated with OD is its assumed potential to provide people with opportunities for innovation by engaging in open processes. Reilly and Smith (2013) argue that social change itself depends on "who can innovate, who benefits from those innovations, and how potentially disruptive the changes are" (p. 39). Here innovation is understood as the result not of individual but of collective efforts. It reflects an argument by von Hippel (2005) that distributing access to 'innovation resources' (e.g. digital tools, information, knowledge) helps "democratise the opportunity to create" (p. 123). Yet little research exists into how innovation manifests through OD. Regarding peer production as an open practice, Benkler (2017b) describes open knowledge flows and sharing as crucial sources for innovation. He finds that peer production networks in the public domain (e.g. information commons) have been particularly effective in distributing knowledge as the basis for innovation activities.

Apart from this, the discourse around innovation in development tends to borrow from the private sector where innovation – as discovery, development and commercialisation of new ideas – is seen as crucial for successful businesses (O'Connor et al., 2018). A popular business method to promote innovation also used in development initiatives,

including the ORI pilot, are hackathons. These are time-bound events in which organisations task participants with solving pre-defined challenges using a set of provided resources (e.g. digital platforms, soft- or hardware). Lilly (2019) finds that these provisions limit the extent to which participants can innovate. Hackathons may help build participants' skills, but they "immerse [them] in the problem-framings offered by [...] institution[s]" (p. 243). Moreover, Zukin and Papadantonakis (2017) find that hackathon organisers benefit the most by outsourcing work, crowdsourcing innovation and increasing their reputation.

This illustrates that assumptions about OD and innovation need to be critically questioned, particularly from a global South perspective. Investigating hackerspaces and open-source projects, Braybrooke and Jordan (2017) discover that the narratives concerning societal benefits of these arguably innovative practices are dominated by three 'technomyths': technological determinism, neoliberal capitalism and Western centrism. Hence, they call for channelling ICTs "towards a greater focus on communal and subaltern uses of technologies outside of dominant Western narratives" (p. 41). Innovation in the global South manifests in unique forms and processes which global Northern discourses of innovation as novelty, progress and growth fail to grasp (Jackson, 2014). In Sub-Saharan Africa, Mbembe (2017) argues, innovation must be understood as deeply embedded in historical contexts affected by colonialism and exploitation. Therefore, the "apparently inexhaustible capacity for creativity and resilience" (n.p.) may only be harnessed locally and can be neither instructed from outside nor simply facilitated by technology.

Yet, ICT4D's tendency to foreground the role of technology in development while neglecting social, economic or political conditions (Murphy & Carmody, 2015; Pieterse, 2010; Unwin, 2017) also extends to OD. Singh and Gurumurthy (2013) find OD initiatives often overlook how socio-economic power structures are replicated within ICT-facilitated networks. Open practices might undermine equity and social justice by "ascribing choices to marginalised communities, subject to deep structural disadvantages, which they simply may not have, and exhorting them to take risks that they may not be able to afford" (p. 185). Realising developmental potentials of open processes in the global South, therefore, also depends – as Graham and Haarstad (2011) argue – on "access being conceived of as embedded in broader processes of development 'on the ground': local capacitation, building of infrastructure, democratisation, and social change" (p. 14).

To conclude, the assumption that ICT-enabled openness decentralises socio-economic power is far from a certain outcome of OD initiatives. Schrape (2019) argues that ICTs may only enhance already existing tendencies towards decentralisation, and that even this will materialise "gradually, through complex processes of deliberate adaption and negotiation" (p. 20). At best, open processes are one of many means to support

development. And while proponents of OD stress its potential for facilitating “a diversity of co-created development paths” (Reilly & Smith, 2013, p. 11), realising this potential requires a thorough understanding, concern for and incorporation of local perspectives, especially regarding how and why actors might be positively or negatively affected by open processes (Bentley 2017).

2.3 Relevance to the ComDev research field

The relevance of my thesis’ topic to communication for development (ComDev) may not be immediately obvious. After all, crowdsourcing an open Kinyarwanda voice dataset does not explicitly fit within common definitions of ComDev as the “strategic communication toward and about social change” (Wilkins, 2015). Viewed from an OD perspective, however, my topic’s more implicit relation to core issues of ComDev becomes apparent.

As outlined, OD is concerned with how ICT-facilitated networks afford people opportunities to collaboratively produce, share and use information in ways that may help improve their lives. Hence, communicative processes are very much at the heart of OD initiatives which connect people, technology and content. Viewing the ORI pilot from this angle shows how it involves actors with varying roles, responsibilities and intentions in crowdsourcing a Kinyarwanda dataset. The analysis will demonstrate how this process is technically centralised via Mozilla’s CV platform and socially decentralised via peer-facilitated, voluntary data contributions in Rwanda. This ‘datafication infrastructure’ raises several questions: In how far is it guided by Rwandan rather than GIZ’s or Mozilla’s objectives? Who is able to utilise the data – or make them usable to others? How are they assumed to yield developmental benefits – and who negotiates what those are in the first place?

Hence, my thesis is positioned between two competing paradigms in the ComDev field (Morris, 2005; Tufte, 2017): the ‘diffusion-of-innovations paradigm’ which regards lacking knowledge and information as problems to be resolved through their top-down provision; and the ‘participatory paradigm’ which aims at empowering citizens to identify problems themselves and resolve them through their own solution strategies. The ORI pilot assumes that creating and providing open access to a Kinyarwanda dataset as a previously non-existing, valuable resource may help promote the capacity to innovate and the co-creation of socially beneficial services or products. At the same time, it envisions the stewardship of this open resource through a community of actors with varying responsibilities for its provision, distribution and usage. Exploring how these assumptions by actors involved in the ORI pilot influence its implementation, the relations between them and underlying development paradigms is the purpose of the following analysis.

3 Methodology

This chapter introduces and outlines the methodology which I have applied for answering my research questions. It first presents the analytical framework I use to elaborate a case study of the ORI pilot – informed by OD and its understanding of openness as social praxis. Then follows an overview of the qualitative data collection methods and scope of my research, as well as a reflection on challenges which I faced.

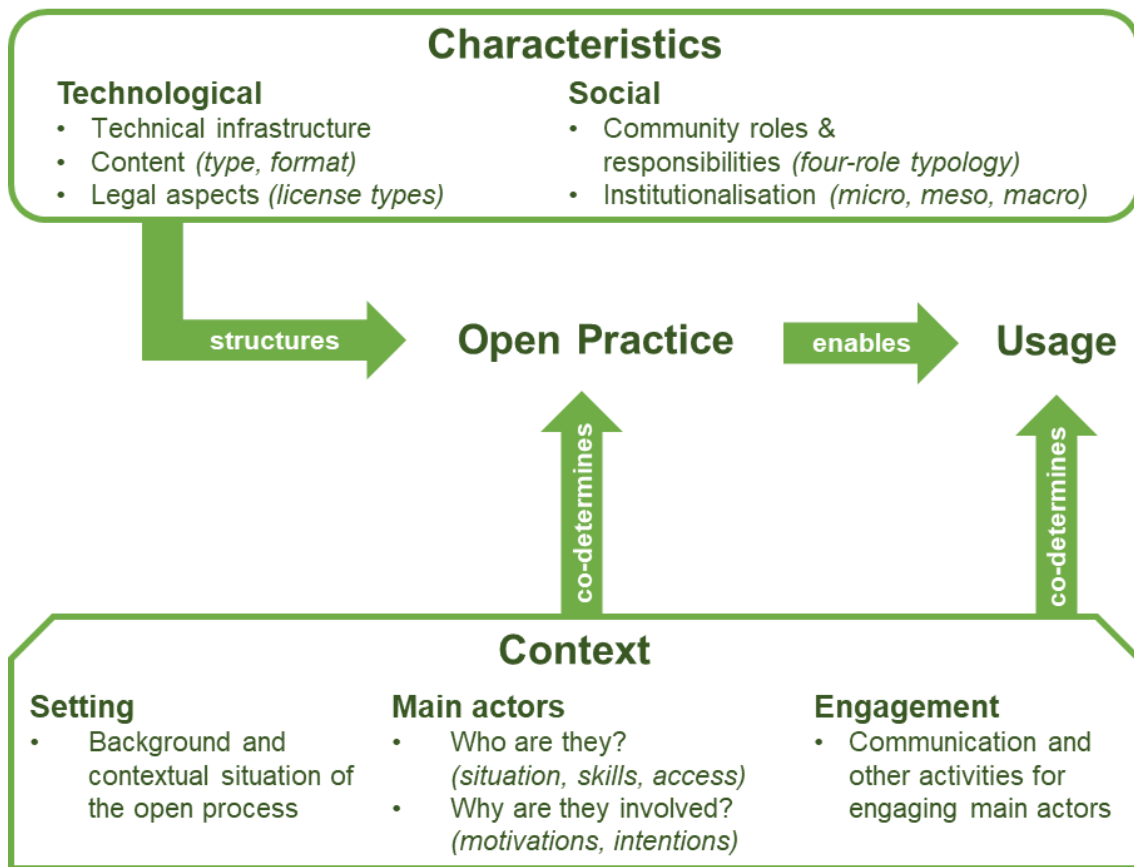
3.1 Analytical framework

To structure my research findings, I use an analytical framework that builds on a model proposed by Smith and Seward (2017) for exploring how open processes – i.e. producing, distributing and using open content – unfold in different contexts and along different open practices, such as crowdsourcing. The model differentiates between four elements of open processes (visualised in Figure 1) which I have adjusted as indicated:

- Firstly, the **context** in which the open process takes place “co-determines how, to what extent and who participates” (n.p.) in it. This covers the overall setting (i.e. background and contextual situation) of the open process and the main actors involved in it, who they are (e.g. their situation, skills or ICT access), why they are involved (e.g. motivations or intentions) and their modes of engagement (e.g. communication or other activities).
- Secondly, the **technological and social characteristics** structure how the open process unfolds. Technologically, this includes its technical infrastructure (e.g. a digital platform), the type and format of the produced content (e.g. voice data) and legal aspects (e.g. copyleft licenses). Socially, it relates to how and by whom the open process is managed. Since Smith and Seward’s model only refers to process ‘drivers’, I added two aspects to also explore the process’s potential for sustainable management: the ‘four-role typology’ of actors to map required community roles and responsibilities (Nickolls, 2017); and ‘institutionalisation’ to assess its anchoring at micro-local, meso-social and macro-institutional level (Singh & Gurumurthy, 2013).
- Thirdly, the **open practice** propels the open process. Smith and Seward identify different practices such as peer production, crowdsourcing, sharing or (re)using content (see chapter 2.2.1). At the heart of the ORI pilot lies the crowdsourcing of Kinyarwanda voice data (see chapter 4.3).
- And fourthly, the **usage** of the open artefact is the outcome of the open practice. As indicated before, I call it ‘usage’ instead of ‘consumption’ to stress the users’

active role. Moreover, Smith and Seward (2017) argue that how open content is used determines how people benefit from it – either directly or indirectly.

Figure 1: Four elements of an open process adapted from Smith and Seward (2017)



For the analysis, I use these four elements to elaborate a case study of the ORI pilot. In qualitative research, case studies are concerned with explanatory descriptions and in-depth understandings of social phenomena, including the discourses, processes and perceptions that frame them (Blatter, 2008). This makes a case study a particularly useful method for me to analyse the ORI pilot and its underlying ‘social phenomenon’ of crowdsourcing an open Kinyarwanda voice dataset as an open process through the lens of OD. Breaking it down into context, characteristics, open practice and usage allows me to explore each element for insights regarding assumptions by and power relations between the involved actors.

3.2 Qualitative methods and scope

To elaborate the case study and investigate my research questions, I collected data using three qualitative methods: participant observation, semi-structured interviews and a focus group discussion.

Through participant observation as part of the ORI project team since February 2019, I gathered the main body of qualitative data for my research. Being involved conceptually and practically in the project, my position proved valuable for three reasons: firstly, to systematically gather 'inside' information; secondly, to critically interrogate it through an 'external' lens; and thirdly, to feed my findings back into the project. As team member and researcher, I took the role of an "observer-as-participant (more participant than observer)" (Gold cited in McKechnie, 2008). This meant that I took notes during weekly 'jour fixes', several in-person workshops of the ORI project team and on-site visits in Kigali. For the analysis, I reviewed all notes and structured my findings along the above-mentioned four dimensions of an open process.

Semi-structured interviews with main actors involved in the ORI pilot helped me gather further insights into their assumptions regarding the developmental impact of the open voice data for Rwanda and the relations between them. Rubin and Rubin (2005) stress that interviewees should be experienced and knowledgeable in the research topic, and should reflect a variety of perspectives. I tried to ensure such a complementarity of perspectives by interviewing actors with different roles in the open process: the start-up *Digital Umuganda* (DU) as the main driver of the voice data collection events, a group of students whom DU trained as facilitators for these events, the manager of the GIZ *Digital Solutions for Sustainable Development* programme in Rwanda which supports DU and the lead for strategic partnerships at Mozilla whose *Common Voice* platform provides the technological crowdsourcing infrastructure.

Finally, I conducted two focus group discussions with the ORI team and DU respectively to critically discuss my main insights pertaining to each research question. This helped me validate and further sharpen my findings, while also reflecting on the ORI pilot's implementation as such. Points of contention concerning certain findings – e.g. regarding underlying development paradigms – as well as my suggestions to improve ORI's community-based stewardship approach will be discussed further in the future.

Regarding the scope of my thesis, two aspects are important to highlight. Firstly, ORI as a service and its Rwanda pilot form an emergent project which was being conceptualised and implemented whilst I was doing my research. Hence, I could not assess the outcome or impact of ORI, but instead opted to explore assumptions, power relations and development paradigms framing it since these are likely to influence its outcomes. This means, however, that my findings are necessarily exploratory in nature.

Secondly, my thesis does not aim to produce generalisable theory about OD. Such an inductive undertaking would require a larger sample size beyond my one case study as

well as more extensive qualitative research, particularly in Rwanda. Instead I rather see the value of my thesis in identifying tendencies which ideally serve two goals: firstly, to help improve the implementation of the ORI pilot (as indicated above); and secondly, to enrich the existing body of knowledge on OD and open processes with another practical example. Obviously, any hypotheses built on these identified tendencies may be either contested or complemented through future research.

3.3 Reflection on challenges

The outlined methodology helped me elaborate what I believe to be an informative case study of the ORI pilot yielding substantial insights regarding my research questions. However, I faced two challenges during my research concerning my role as observing participant and the scope of my research in Rwanda.

The first challenge relates to my role as observing participant which required conscious reflexivity about my double-position in the ORI project team. I noticed early-on that I had to be careful not to impose my own views or expectations from a research perspective onto the ORI project. At times, it became difficult to distinguish between my operational role as participant in the project (=creatively involved) and my analytical role as researcher of the project (=critically observing). This led me to take care not to conflate team insights with my own by keeping separate notes during our team engagements. Moreover, I tried to heed the advice by Bentley et al (2019) that researchers must not impose their own normative ideals regarding openness onto OD initiatives at the risk of neglecting local perspectives. This meant juxtaposing my analytical independence against my vested interest in ORI. Asking myself how I could critically observe a project which I obviously want to succeed, I found my academically informed perspective and critical reflections to be useful for informing ORI conceptually and in practice.

The second challenge I faced relates to the limited scope of research I was able to conduct in Rwanda. Apart from two visits to Kigali with the ORI team, I observed the open practice of crowdsourcing voice data at a distance, mediated via insights shared by the start-up *Digital Umuganda* which is organising the data collection events. More research among participants of these events would have been ideal for two reasons: firstly, learning about the impacts of datafication on communities in the global South requires taking these communities as the starting point of engaged research (Milan & Treré, 2019); and secondly, thorough power analysis requires close work with communities living enmeshed in those power relations (McGee & Pettit, 2019). Although I have done the best I could with the time and resources at hand, I need to acknowledge that the insights I gathered in Rwanda may only indicate certain tendencies that require more in-depth research for better understanding them.

4 Analysis: The GIZ Open Resources Incubator pilot for crowdsourcing open voice data in Rwanda

“With voice interaction available in their own language we may provide millions of people access to information and ultimately make technology more inclusive.”

Alex Klepel (Mozilla) & Lea Gimpel (GIZ) (2019)

“We are bringing Kinyarwanda to the digital age.”

Student volunteers in Kigali

Since its launch in February 2019, the ORI pilot has unfolded as an open process aimed at crowdsourcing a Kinyarwanda voice dataset whilst establishing a community of actors for its sustained management and stewardship. The following analysis investigates the ORI pilot in the form of a case study crafted along the four elements of an open process: firstly, it outlines its dual context as a GIZ idea piloted in Rwanda to test assumptions about voice data; secondly, it illustrates its characteristics as a socio-technical process involving people and technology; thirdly, it presents the crowdsourcing as a global-local practice that is globally centralised and locally decentralised; and fourthly, it explores considerations regarding the usage of the open voice dataset and the chains of intermediaries it requires for yielding societal benefits.

Throughout the analysis, findings pertaining to assumptions which the involved actors hold regarding developmental benefits of open voice data will be highlighted (= main research question). The introductory quotes exemplify the range of such assumptions – from better information access and more inclusive technology to digital representations of Kinyarwanda. Moreover, power relations manifesting between the involved actors (= first sub-question) will also be outlined. Finally, underlying development paradigms (= second sub-question) will be addressed as part of discussing the main findings in the conclusion.

4.1 Dual context as GIZ idea piloted in Rwanda

The ORI pilot takes place in Rwanda where it established a community of actors in the crowdsourcing of an open Kinyarwanda dataset. That is its local context. The idea for this did not originate in Rwanda, however, but was introduced by GIZ and Mozilla. That is its international. To what extent the implementation of the ORI pilot has been determined by this ‘dual context’ will be analysed in this chapter.

The idea behind the ORI pilot emerged from the 2018/19 round of the GIZ *Innovation Fund*, a company-wide programme for “developing, implementing and rolling out innovative ideas for products, services and business models” (GIZ, n.d.-b). Through this programme, a team of two GIZ senior planning officers in Germany, a GIZ development advisor in Rwanda and me as external consultant won seed funding to explore how voice-recognition

technology for languages in the global South could be made openly available. Our team assumed that such open access could enable local actors to ‘develop innovative local voice-based solutions’ e.g. simplifying access to social or health-related information services (Gimpel et al., 2019a).

Progressing from this idea, our team conceptualised the *Open Resources Incubator* (ORI) as a service to facilitate the collective production, distribution and use of openly available content – such as open voice data in Rwanda. At its core lies the ambition to establish the community-based stewardship of such an open resource by identifying, connecting and supporting key actors with shared responsibilities for its sustained management. ‘Community’ is understood not as locally bound but as a broad, even international network of actors cooperating in “partnerships of equals, truly empowering for all parties engaged.” Moreover, open resources are framed as potential “catalyst[s] for sustainable development” which may “lay the ground for a more equal distribution of innovation capacities” and “contribute to levelling the playing field by enabling open access to information, technology and the means for local value creation, while avoiding the monopolisation of economic power over digital technologies and knowledge” (Gimpel et al., 2019b).

The Mozilla Foundation became the main international partner for the ORI pilot. Their *Common Voice* (CV) project provides an open online platform for crowdsourcing voice data in any language. It is backed by machine-learning (ML) experts who develop openly accessible speech models based on the collected voice data. With CV, Mozilla aims to address biases in voice-recognition technology such as lacking accessibility beyond large corporations or lacking representativeness beyond global majority languages. The project calls on volunteers to “donate [their] voice to help us build an open-source voice database that anyone can use to make innovative apps for devices and the web” (Mozilla, n.d.-b). Mozilla values cooperating with GIZ as an opportunity to extend CV to Sub-Saharan languages not yet digitally represented while contributing to sustainable development through locally driven technological innovation. Mozilla shares the ORI team’s assumption about the innovative potential of voice-interaction to simplify people’s access to information and make technology more inclusive (Klepel & Gimpel, 2019).

Rwanda as the local context for the pilot was deliberately chosen by the ORI team which saw its emerging ICT sector and ICT-focused public policy as a promising environment to test its assumptions regarding open voice data. Moreover, GIZ’s Kigali-based *Digital Solutions for Sustainable Development* (DSSD) programme has been supportive of the ORI pilot which its manager describes as laying ‘technical groundwork’ to open a new technology to Rwanda. This fits to DSSD’s goal of “developing and implementing SDG-related [...] digital solutions” (GIZ, n.d.-a) jointly with Rwandan public and private partners.

The pilot kicked off in February 2019 with a hackathon co-organised by the ORI team and Mozilla at the *kLab* innovation hub in Kigali (see Figure 2). Participating entrepreneurs were tasked with finding creative means for collecting a speech corpus in Kinyarwanda using the CV platform (Klepel & Gimpel, 2019). The event also managed to raise interest among Rwandan public and private actors in exploring the potential of openly accessible voice data to improve their services and products (see chapter 4.4).

Figure 2: Participants of the GIZ/Mozilla hackathon at *kLab* in Kigali (Photo: D. Brumund)



The winning proposal came from two young Rwandan entrepreneurs with prior start-up experience (see Figure 3): They found that the monthly *Umuganda* community works programme and its ideal of collectively contributing to improving Rwanda could be used to mobilise young Rwandans, such as students, interested in volunteering to record voice data as something more ‘fun’ than usual *Umuganda* activities e.g. cleaning public spaces.

Following the hackathon, this duo formed the start-up *Digital Umuganda* (DU) as an intermediary linking international initiatives such as *Common Voice* with local communities and contexts. DU views open voice data as ‘digital infrastructure’ providing opportunities for local developers to build digital tools with technology they would otherwise not have access to; and for Rwandan citizens who may benefit from improved access to information and services via voice-recognition (Digital Umuganda, n.d.).

Figure 3: Winning team (middle) and founders of start-up *Digital Umuganda* (Photo: D. Brumund)



Since then, DU has become the main actor in Rwanda to foster and coordinate the community of actors involved in crowdsourcing the Kinyarwanda dataset. This includes universities whose students volunteer to record voice data via CV, media companies who provide Kinyarwanda text-data as well as public and private actors who show interest in using voice interaction to improve their products or services (see chapter 4.2). Among these actors, the partner universities are particularly important as the venues they provide and the volunteer work of their students underpin the crowdsourcing practice of producing the Kinyarwanda dataset (see chapter 4.3). The students' motivations for participating in the project range from the wish to support development in Rwanda to pride in 'bringing Kinyarwanda to the digital age', while some even hope to learn how to program their own applications with it.

To conclude, the influence of the ORI pilot's dual context on its implementation in Rwanda is evident. Its international origin as a GIZ-internal idea situated GIZ and Mozilla as main actors, and established CV as digital platform. These actors deliberately chose Rwanda as local context to test their assumptions that open voice data distributes innovation capacities, promote entrepreneurship and create social impact e.g. by improving access to services. The fact that the ORI pilot was not initiated based on Rwandan interests or aspirations raises the question to what extent its implementation is driven by local rather than international objectives. In this regard, providing entrepreneurs with access to novel technologies aligns with Rwandan public policy, particularly the *Smart Rwanda 2020* NDP and its objective to "improve and expand access to ICT skills and innovation capacity" (MYICT, 2015, p. 22). Moreover, the hackathon in Kigali served to identify and convene

Rwandan actors – particularly DU and universities – who adopted the assumptions around Kinyarwanda voice data and began its crowdsourcing practice, thereby initiating the local part of its community-based stewardship.

4.2 Socio-technical process with emergent institutionalisation

Involving people and technology in crowdsourcing the Kinyarwanda dataset, the ORI pilot unfolds as a socio-technical process. Its technological and social characteristics in terms of Mozilla’s CV infrastructure and the voice-data content as well as the community-building among the Rwandan and international actors will be analysed in this chapter.

As shown, the technological characteristics of this process were pre-determined by the partnership between GIZ and Mozilla which established the CV platform as technical infrastructure for the crowdsourcing practice. The platform facilitates two steps: firstly, the recording of voice by reading out pre-defined sentences; and secondly, the verification of voice recordings by listening to and validating them (see Figure 4 and Figure 5). Both steps can be performed by any site visitor without registration. Mozilla encourages users to set up profiles for enriching the dataset with basic demographic data (i.e. gender, age) and for keeping track of their contribution statistics or comparing them with others.

Figure 4: Mozilla’s *Common Voice* platform for voice recordings in Kinyarwanda

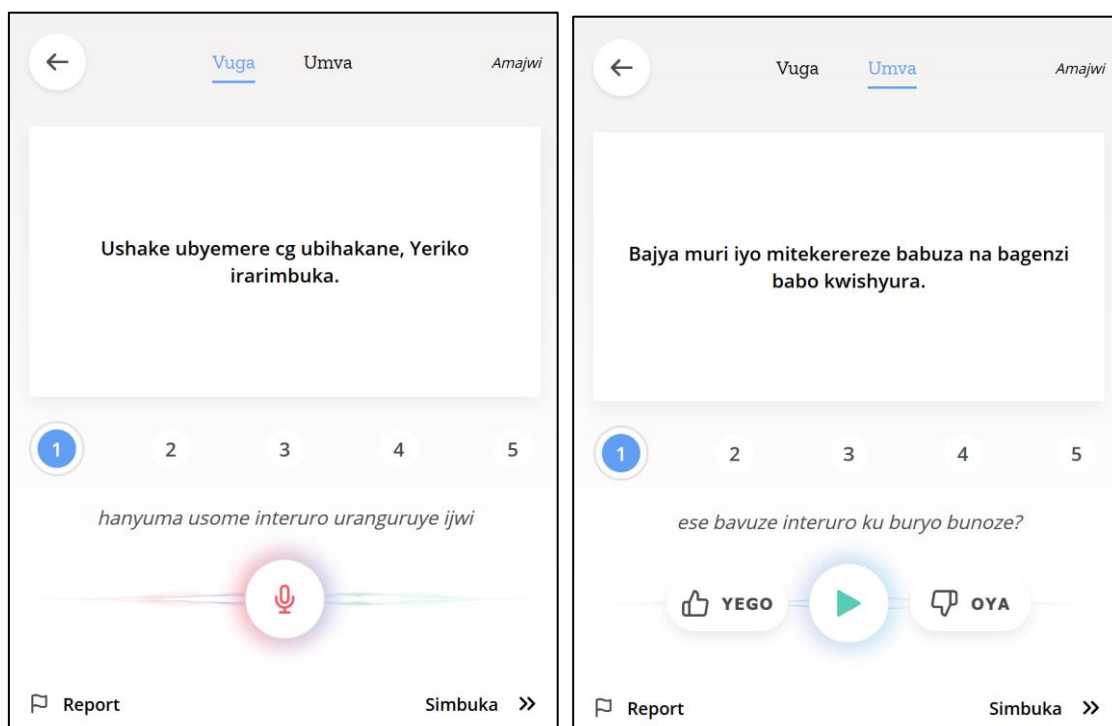
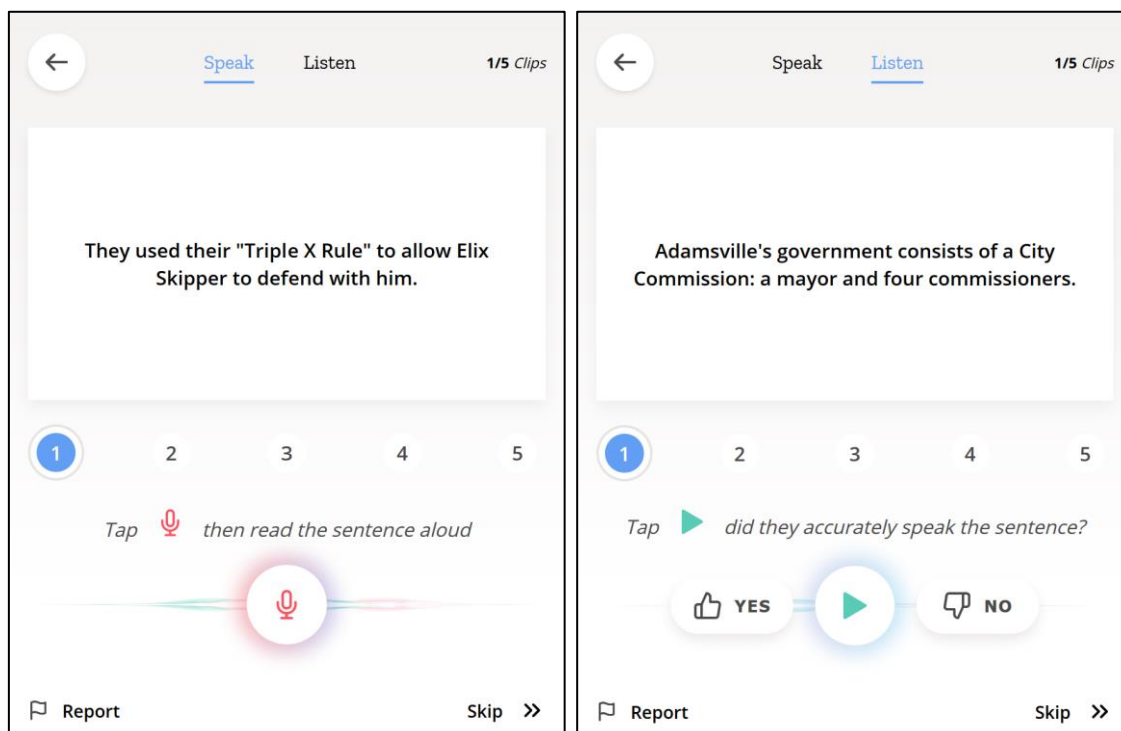


Figure 5: Mozilla's *Common Voice* platform for voice recordings in English (as translation)



Prior to its use in Rwanda, the CV platform had to be 'localised' by translating its static texts and uploading copyright-free sentences in Kinyarwanda as preparation for the voice recordings. For this, Mozilla provides a basic, web-based tool which allows users to upload sentences in the public domain in any pre-defined language and review them for their correctness (see Figure 6). Technical challenges regarding both tools – the main CV platform and the 'sentence collector' – will be addressed in the next chapter.

Figure 6: *Common Voice* 'Sentence collector' for adding and reviewing text data

Add Sentences

Select Language

Kinyarwanda (Ikinyarwanda)

Add public domain sentences

One sentence per line

Where are these public domain sentences from?

Read our How-to if unsure how to attribute

☐ I confirm that these sentences are public domain and I have permission to upload them.

Submit

Review Sentences

English (English) Review Criteria

"Oh, I forgot you," said the tyrant.

Oh

Come, I'm not going to nurse him.

Oh

I snatched it away, however, and seized his arm.

Oh

Finish Review

1 < 1 > 2000

The CV platform provides continuous open access to the crowdsourced voice dataset consisting of the Kinyarwanda texts and all corresponding voice recordings verified by at least two users. Based on these datasets, Mozilla aims to provide openly accessible ‘speech models’ using its open-source voice-recognition engine *DeepSpeech*. To achieve this, different quantities of voice data are required: for speech-to-text models, at least 1.200 hours of voice recordings taken in noisy, real-life environments; and for text-to-speech models, at least 20 hours of voice recordings in good, ideally studio-quality.

There are two main technical challenges concerning speech models. Firstly, crowd-sourcing the required amount of voice data is challenging. Even for English, CV’s most prominent language, a complete dataset is not yet available. Therefore, Mozilla’s ML team is working on small-scale models which recognise limited amounts of words for specific applications (e.g. numbers, or sector-specific terms). That points to the second challenge: a speech model only recognises words present in its underlying dataset. This means either sufficiently large datasets (for broad usage) or specific ones (for targeted use in specific sectors) are needed.

As legal basis for CV, Mozilla uses a CC0 1.0 license which dedicates its content to the public domain without copyright, thereby allowing users to “copy, modify, distribute and perform [it], even for commercial purposes, all without asking permission” (Creative Commons, n.d.). This CC0 license means, firstly, that contributors forfeit the rights to their text and voice contributions and, secondly, that users of the dataset are unrestricted in how they apply it. This has implications for the ORI pilot and how to ensure that the open voice data serve to create social impact – which will be investigated in chapter 4.4.

Moving from the technological to the social characteristics of the ORI pilot, a first aspect to note is how these are mutually dependent: While the CV platform structures how the open process unfolds technically, a community of actors in Rwanda is required to sustain it, to ensure it yields voice data in sufficient quality, and to create benefit from it. An actor mapping along the four-role typology by Nickolls (2017) – maintainers, contributors, users and sustainers – illustrates roles, responsibilities and relations within the community that emerged throughout the ORI pilot.

The maintenance of the open process moved from GIZ and Mozilla, who initiated it, to DU after the hackathon in Kigali who has been coordinating the crowdsourcing of the Kinyarwanda dataset (see chapter 4.3). Technically, DU is dependent on Mozilla’s CV platform as digital infrastructure and on their ML experts for eventually developing a Kinyarwanda speech model. DU is intent, however, to complement such international

dependencies with local co-operations – e.g. with Kigali-based data-science company *Hepta Analytics* – thereby fostering a local community of practice.

There are two groups of contributors to the crowdsourcing practice. Firstly, several media outlets support DU with providing CC0-licensed Kinyarwanda texts needed to set-up the CV platform. Secondly, and most importantly, students volunteer at peer-facilitated data collection events which are coordinated by DU. Partner universities provide facilities as well as devices such as laptops for these events. The crowdsourcing practice will be explored further in chapter 4.3.

In terms of users, a growing number of public and private actors show interest in utilising voice interaction for improving their products or services (see chapter 4.4). DU engages them through bilateral meetings and ‘voice-community meetups’ which further contribute to the aforementioned community of practice whilst raising awareness of the potentials of voice interaction for sustainable development (Digital Umuganda & GIZ, 2019).

Finally, the open process is sustained through support from GIZ: financially, by the DSSD programme which funds DU to coordinate the collection of 1.200 hours Kinyarwanda voice data; and operationally, by the ORI team which supports DU with organisational and business development as well as networking opportunities (Digital Umuganda & GIZ, 2019).

The mapping reveals a community which is not (yet) self-reliant but dependent on hierarchical relations between international and local actors (Mozilla-GIZ-DU), and among local actors (DU-Universities-Students) for driving the crowdsourcing practice. Moreover, the community is barely integrated: The different actors are connected via DU but without direct engagement e.g. between the student contributors, Mozilla or any of the potential public/private users. This means that, as of yet, the stewardship of the open process is not community-based (as envisioned) but rather dominated by DU and Mozilla, as well as dependent on GIZ for financial support.

These dominant actors recognise the need for closer community integration towards what Mozilla describes as ‘self-sustaining cycles’ of stewardship. Sustaining the process beyond the pilot phase will require them to address three challenges: firstly, alternative, long-term funding sources to replace its donor support are not yet in sight; secondly, the role distribution among actors with different means to benefit from the process complicates community-building on equal terms (e.g. students contributors vs. public/private users); and thirdly, the process is highly dependent on DU as ‘nodal point’ in Rwanda which raises the question if additional actors like them are needed to help maintain, sustain and upscale it.

These challenges show that the extent to which the community around the open process will be institutionally anchored in Rwanda – at micro-social, meso-social and macro-institutional

level (Singh & Gurumurthy, 2013) – may be of prime importance for its sustainability. In this regard, agreements DU is entering into with public, private and academic actors to support the crowdsourcing process point to an increasing institutionalisation at meso-social level. This could be expanded to micro-local level, e.g. by private or academic actors capacitating local developers to use speech models, and to macro-institutional level, e.g. through public policy support for community-based stewardship of open resources.

To conclude, the ORI pilot established a ‘socio-technical’ process that is characterised technologically by Mozilla’s CV platform as well as socially by a community of international and national actors intended to take on its stewardship. Due to its hierarchical set-up and the actors’ varying degrees of influence, this community is currently barely integrated, rendering the process dependent on few dominant actors, particularly GIZ, Mozilla and DU. Yet, DU’s efforts to foster a local community of practice indicate an emergent institutional anchoring important for the process’s sustainability.

4.3 Crowdsourcing as global-local open practice

The crowdsourcing of the Kinyarwanda dataset is organised on two levels: globally via Mozilla’s CV platform as its technological infrastructure, and locally via DU as central organiser and co-ordinator of data collection events in Rwanda. Hence, the crowdsourcing takes the form of a global-local open practice aimed at producing the two parts of the dataset: Kinyarwanda texts and corresponding voice recordings. How this global-local practice unfolds throughout the ORI pilot will be analysed in this chapter.

As prerequisite for the voice recordings, a corpus of at least 5.000 copyright-free sentences in Kinyarwanda had to be uploaded and reviewed on the CV platform. Finding suitable text sources was a challenge due to licensing issues. For instance, the use of bible texts – which DU assumed to be in the public domain – failed because the translations were copyrighted. Eventually DU set up an agreement with Rwandan online media company IGIHE which committed to releasing up to 1.2 million Kinyarwanda sentences from their archives to the public domain. Mozilla supported the contractual set-up with legal advice on CC0-licensing which was new to IGIHE – indicating a need for IP training similar to what De Beer and Ouagamanam (2013) identified in Nigeria. Uploading and reviewing these sentences on the CV platform became part of the data collection events further described below.

Building on their hackathon idea, DU has been organising and co-ordinating the crowdsourcing as peer-facilitated data collection events at universities. They formalised cooperation agreements with over 30 universities which provide venues and devices (e.g. laptops) for the events. At each university, DU trains peer facilitators in using the

CV platform, mobilising students and conducting data collection events. The facilitators are called 'commoneers': *Common Voice* pioneers who contribute to a common, digital resource for Rwanda. They receive basic monetary reimbursement for travel- and logistics-related costs. Through the commoneers, DU aims to distribute and upscale the crowd-sourcing of the Kinyarwanda dataset. Their goal is to produce the 1.200 hours of validated data recordings needed for a Kinyarwanda speech-to-text model within 12 months – a timeframe they committed themselves to vis-à-vis GIZ's DSSD programme which provides financial support.

Prior to each data collection event, commoneers raise awareness at their universities about the potential of open voice data for Rwanda. One commoneer explained how they mobilise students by appealing to their wish to contribute to Rwanda's development and presenting voice data as something 'made by them and for them'. DU stresses further that what excites the students is the idea of digitising Kinyarwanda.

The events are limited to 30-40 students and follow a half-day programme: The students are first introduced to the CV platform, mostly using laptops provided by the university. They then continue to participate in two blocks of tasks: firstly, uploading and reviewing Kinyarwanda sentences; and secondly, adding and validating voice recordings (see Figure 7). At the end, they receive a participation certificate issued by Mozilla stating their name and contributions.

Figure 7: Data collection events at universities organised by Digital Umuganda (Photos: DU)





Following a first round of university events, DU found it eye-opening to realise how difficult the voice data collection was due to technical and content-related challenges. At many universities, unstable internet networks complicated using the CV platform which obliged DU to invest in mobile WiFi routers as a temporary solution. Moreover, some aspects of the CV platform itself were challenging – particularly the sentence collector which required sentences to be uploaded individually. These issues were directly addressed by Mozilla adding a feature for bulk uploading texts and easier review mechanisms. Furthermore, it proved challenging to produce voice data in sufficient quantity and quality. Within the first three months of data collection, only 17 hours of voice data were recorded of which 17 minutes were validated (Mozilla, n.d.-a). There also seems to be a notable gender bias with 41% of contributors being male and only 8% female – though 50% are unspecified.

In view of this, it may be necessary to consider alternative technical and/or social methods for upscaling the crowdsourcing practice as planned. For instance, Mozilla plans to add a reversed data collection mechanism starting with uploading audio recordings and then crowdsourcing their transcription. Regarding social methods, DU has extended the data collection events to collaborating high schools, and on-boarded student supervisors and teachers as communeers to facilitate the events. Moreover, it may be worthwhile to investigate the mobilising appeal of *Umuganda*. DU views it as a ‘communal approach of self-care and voluntary contributions’, yet it is also politically instructed and socially monitored (Purdeková, 2016). In-depth research among participants could help explore to what extent *Umuganda* inspires or compels them to join the data collection events.

To conclude, the crowdsourcing practice at the core of the ORI pilot takes the form of a global-local datafication infrastructure which is globally centralised via Mozilla's CV platform and locally decentralised via distributed, peer-facilitated data collection events mobilising students for voluntary contributions. If this arrangement can scale sufficiently to yield 1.200 hours of validated Kinyarwanda voice recordings in twelve months remains to be seen. It will depend on at least two aspects: firstly, on addressing the challenges concerning data quality and quantity, e.g. with improved technical and social crowdsourcing methods; and secondly, on sustaining the students' motivation which hinges on whether they continue to see benefit in participating, both for themselves and for Rwanda.

4.4 Intermediated usage and indirect developmental benefits

Since the crowdsourcing of the Kinyarwanda dataset is on-going and a speech model not yet available, any considerations regarding their usage by actors involved in the ORI pilot are largely hypothetical. Yet, the ways the data will be used ultimately determine how people benefit from them (Smith & Seward, 2017). Therefore, this chapter will analyse both the usability of the open voice data and their potential for development.

A first, important aspect to note is that the Kinyarwanda dataset and speech model will not be usable to everyone in the same way. Improving access to public services via voice interaction, for instance, presupposes that a public institution is willing to invest in such an improvement and that it contracts experts in ML and IT development. This illustrates that a series of intermediary steps is required before the open Kinyarwanda dataset may yield any benefits, whether social or commercial, and that there is a difference between those who can benefit directly or only indirectly from it.

Within the community of actors established through the ORI pilot, those engaged as maintainers and users are most likely to benefit directly from using the voice dataset. On the one hand, there are Mozilla, the Rwandan data science firm *Hepta Analytics* and DU itself who plans to employ an IT developer. With their ML experts skilled in training speech models or their IT developers skilled in developing applications with those models, these actors can directly utilise the open voice dataset – and make it usable to others. Hence, they can be considered first-level intermediaries.

On the other hand, there are public and private actors interested in using voice interaction to improve their products or services for citizens and consumers. From the public sector, they include institutions such as RISA and RURA with whom DU signed agreements e.g. to support improving national services such as the eGovernment portal *Irembo*. The National Council for Persons with Disabilities also expressed interest in applications for reading out digital information such as news to sight- or speech-impaired citizens.

And from the private sector, they include Rwandan smartphone manufacturer *MaraPhone* who is interested in offering devices off-the-shelf with Kinyarwanda voice-recognition, or UK-based digital health information provider *Babyl* who offers SMS/USSD-transmitted services in Rwanda and is interested in simplifying access via voice commands. These public and private actors require ML experts or IT developers to realise the envisioned improvements and can, therefore, be considered second-level intermediaries.

It is through these intermediaries that Rwandan citizens may benefit from the Kinyarwanda dataset, albeit indirectly by using improved services or purchasing new products offered by those public and private actors. Most notably, also students who volunteer at data collection events stand to benefit only indirectly from the output which they produce. As contributors, they are a crucial part of the community which is to take on the stewardship of the open Kinyarwanda dataset. Yet, they have no influence over what products or services might be developed with it, or whether these yield any benefits for themselves, or for Rwanda.

This has important implications for the ORI pilot: As shown, a main motivation for students to participate in the crowdsourcing practice is their wish to do something good for Rwanda. If they volunteer assuming that their contributions will benefit the country at large, then it ought to be the ORI pilot's obligation to ensure that this assumption is realised. Yet, the fact that the Kinyarwanda dataset and future speech models are dedicated to the public domain may pose a challenge: It leaves it to the (intermediating) data users' discretion to create products and service that yield societal benefits, or solely profit-oriented ones.

Insights from the ORI pilot point to two strategies for addressing this challenge. Firstly, the identification and realisation of socially beneficial use cases is promoted by DU who consults public institutions in this regard and advocates for supporting impact-driven NGOs with technical skills to benefit from voice interaction, as well as by GIZ and Mozilla who expanded their cooperation as part of a new AI-focused initiative which promotes 'locally suitable, voice-enabled products or technologies' relevant to the SDGs – initially focusing on start-ups in Rwanda (Mozilla, 2019). And secondly, both GIZ and DU argue that the stewardship of the Kinyarwanda dataset should balance non-profit and for-profit usage. For instance, sustaining the crowdsourcing effort beyond donor support requires alternative funding sources which could come from private actors. Their support is more likely if commercial usage of open voice data is first encouraged and then linked to a commitment to re-invest certain profits back into the voice data community.

To conclude, the fact that the open voice data requires chains of intermediaries to be made usable, both for social and commercial purposes, means that also developmental benefits may only be achieved indirectly – at the discretion of these intermediaries. Ideally, public

and private actors create products or services that go beyond profit motives to also serve society at large, e.g. improving access to information or simplifying the use of technology. Yet, the unrestricted openness of the data lacks provisions to guide such socially beneficial usage. Although GIZ, Mozilla and DU are promoting the realisation of SDG-relevant use case of open voice data in Rwanda, the question remains how this might be sustained, e.g. as part of the community-based stewardship, to balance non-profit/for-profit usage and curtail market dominance.

5 Conclusion

“Appropriate global governance of the digital should promote national and local digital economies. It ought to ensure that competitive and open global technical services are accessible locally – including by local digital businesses – on fair and regulated terms. [...] A new digital model that is local-to-global must be shaped, which supports localness and furthers democratic self-determination, without compromising on the important benefits of the globalness of the digital.”

‘Digital Justice Manifesto’ by the Just Net Coalition (2019)

As of February 2020, the GIZ *Open Resources Incubator*’s pilot project in Rwanda remains at an early stage: The crowdsourcing of a Kinyarwanda dataset sufficiently large to train a speech model is on-going, as are efforts to establish a self-sustaining community of actors for its stewardship. Although outcomes of the ORI pilot can only be assessed once the dataset or speech model start being used, my thesis explored three aspects that already influence its implementation and potential impact: firstly, assumptions about developmental benefits that the involved actors associate with open voice data; secondly, power relations manifesting between these actors; and thirdly, any underlying development paradigms.

The analysis showed how the ORI pilot is defined by a dual context: It originated as a GIZ-internal idea in cooperation with Mozilla whose CV platform facilitates the global crowdsourcing of open voice datasets. Both GIZ and Mozilla assume that open access to voice data can help distribute innovation capacities, promote entrepreneurship and yield social impact e.g. by improving access to services or making technology more inclusive.

As local context to test the validity of these universal assumptions in the global South, they chose Rwanda for its ICT-friendly policy environment and a supportive GIZ network on-site. A hackathon in Kigali served to identify and connect Rwandan actors who shared these assumptions. These include public and private institutions intrigued by the potential of voice technology to improve their services or products, as well as particularly the start-up *Digital Umuganda* which was founded by the hackathon’s winning team. Through cooperation with universities who provide facilities and trainings of peer facilitators who mobilise students for data collection events, DU began coordinating the crowdsourcing of

the Kinyarwanda dataset and initiating the local part of its community-based stewardship. The fact that these actors, by and large, share the same basic assumptions about benefits of open voice data indicates that it might be too abstract to discuss the potentials of such data without yet having realised any use cases exemplifying this potential in practice.

More interestingly, the analysis yielded important findings regarding the relations between the different actors involved in the ORI pilot. Their relations appear to have more significant influence on the pilot's implementation than their assumptions – counter to what the main research question had implied. These relations manifest on two levels which structure how the pilot unfolds and co-determine who benefits from the open Kinyarwanda dataset.

On a first level, the ORI pilot established a socio-technical process that connects Mozilla's *Common Voice* platform with a community of international and local actors who are involved in producing the open voice dataset and intended to take on its stewardship. An actor mapping revealed hierarchical relations between the international and local actors, and among the local actors. A closer, more balanced integration of this community is necessary but complicated by the dominant positions DU (locally) and Mozilla (internationally), the financial dependence on GIZ as well as these actors' varying degrees of influence over how and for whose benefit the process unfolds. However, DU's efforts to foster a local community of practice – through cooperation with public, private and academic institutions – indicate an emergent institutional anchoring in Rwanda which should be expanded for the process's long-term sustainability.

On a second level, the crowdsourcing practice itself manifests as a global-local datafication infrastructure that sources voice data from Rwandan volunteers for copyright-free use. It is globally centralised via Mozilla's CV platform and locally decentralised via peer-facilitated data collection events which DU coordinates to convene students at partner universities. The analysis showed that upscaling this practice to produce Kinyarwanda voice data in the quantity and quality needed for training a speech model may require improvements of the CV platform's technical features as well as of the social mechanisms for mobilising students (and others) as volunteers. As one of their main reason to participate is the wish to do something good for Rwanda, the students' motivations might only be sustained if they see the open voice data being used for socially beneficial purposes.

In this regard, the analysis outlined how chains of intermediaries – including data scientists, IT developers and funders – are required to make the voice data usable, for either social or commercial purposes. Yet without provisions that guide socially-oriented rather than purely profit-focused uses of the Kinyarwanda dataset, its assumed developmental benefits will only be achieved indirectly – and at the discretion of the intermediaries. In this way,

the current global-local datafication infrastructure establishes a division of labour between public/private actors who have the means to create value and thereby benefit directly from the voice data, and students (or other Rwandans) who voluntarily produce voice data but may, at best, benefit indirectly from it. The contributors' lack of control over how their donated data are used may lead to an exploitative relationship (a risk stressed by Couldry & Mejias, 2019a, 2019c): The analysis showed how students are mobilised by the appeal that their voice data donations will contribute to Rwanda's progress – but how the data usage for social (and not just commercial) impact will be sustained remains unclear.

Hence, the question remains what this tells us about development paradigms underlying the ORI pilot in Rwanda. As shown, its basic premise is that open voice data contribute to development by affording previously inaccessible opportunities for innovation and local value creation. But what type of development – and for whose benefit will the created value accrue? The analysis suggests two development paradigms either of which the ORI pilot might serve depending on how the voice data usage will be directed and institutionalised.

On the one hand, coupling the open access to the Kinyarwanda dataset with an emphasis on entrepreneurship (e.g. by mainly supporting start-ups to use the data) would promote market-driven development whilst favouring those already empowered to utilise open data (resembling a critique by Gurstein, 2011). This would serve a neoliberal paradigm whose focus on deregulated data access neglects inequalities in data usability whilst risking the exploitation of voluntary data contributors for commercial rather than social value creation.

On the other hand, complementing the open access to the Kinyarwanda dataset with efforts to promote its socially-oriented usage as defined by everyone involved in the process would promote participatory development whilst allowing for negotiations of what 'developmental benefit' might mean to different people. This would serve a capabilities paradigm focused on giving people the opportunity and proactively capacitating them to direct innovation and value creation from the open voice data towards their expectations.

It should come as little surprise that my thesis advocates for the capabilities paradigm which aligns with ORI's aim to facilitate open access and usage of resources such as voice data as a 'catalyst' for sustainable development, and not for privatised rewards. My thesis showed, however, that openness of data is not enough to direct its usage towards social impact. It confirms the general critique levelled against OD that open processes replicate – rather than reduce – existing inequalities and power imbalances. Whereas the open production and distribution of the Kinyarwanda data are technically (CV platform), legally (CC0 license) and socially (peer-facilitated data collection) formalised, its anticipated

open usage is entirely unregulated. As a result, it is left to the intermediaries to decide for what purpose to use it.

In view of this, I conclude my thesis by turning towards ORI's goal to establish a community-based stewardship of the open Kinyarwanda dataset. As shown, this goal is hampered by a lack of integration and hierarchical relations within the community as well as lacking directedness towards social impact. To address this, I argue that an institutionalised set of principles – in form of a 'stewardship agreement' – is required which directs the management and usage of the open voice data on joint terms. The negotiation of such an agreement needs to involve all actors in the community: contributors and users, maintainers and sustainers. Together they should reflect on their reciprocal relationships and define how the value they create with the open voice data benefits their community (e.g. through funding or training measures), and Rwanda at large.

By providing a normative framing, stewardship agreements would essentially recognise recommendations by scholars including Mazzucato (2018a), Morozov (2018a, 2019) or Couldry and Mejias (2019b) that equitable, social use of data requires locally negotiated rules and conditions. Examples for suitable principles from a global South perspective can be found in the 'Digital Justice Manifesto' by the Just Net Coalition (2019) which – as per their introductory quote – calls for local-to-global digital models that are fairly regulated and locally self-determined. It outlines principles such as collective control over datafication infrastructures, regulated use of 'data commons' or data rights for their creators.

The usefulness of such stewardship agreements obviously needs to be assessed through research: How can they be negotiated on equal terms despite power imbalances between the involved actors? How can they be realised as legally non-binding but morally compelling rules between local and global actors? And can they help establish equitable datafication infrastructures between actors with unequal means to benefit from the data? Insights may be found in emergent research on 'data justice' (Dencik et al., 2019; Heeks & Renken, 2018; Heeks & Shekhar, 2019) and 'design justice' (Costanza-Chock, 2018, 2020) which puts marginalised communities at the centre of socio-technical processes to challenge and dismantle – rather than reproduce – social inequalities.

Finally, if proven effective, stewardship agreements might go a long way to realise the 'partnership of equals' which ORI envisions for the community-based stewardship of open resources – thereby benefitting not the privileged few, but the deserving many.

(14.398 words)

References

- Andrejevic, M. (2014). The Big Data Divide. *International Journal of Communication*, 8, 1673–1689.
- Arora, P. (2016). The Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication*, 2016(10), 1681–1699.
- Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., & Weyl, E. G. (2018). Should We Treat Data as Labor? Moving beyond 'Free'. *American Economic Association Papers & Proceedings*, 1(1 (Forthcoming)), 38–42. <https://doi.org/10.1257/pandp.20181003>
- Basaninyenzi, U. (2013, November 29). *Paul Kagame: A digital president leading a technology movement*. How We Made It In Africa. <https://www.howwemadeitinafrica.com/paul-kagame-a-digital-president-leading-a-technology-movement/32959/>
- Bellman, E. (2017, August 7). The End of Typing: The Next Billion Mobile Users Will Rely on Video and Voice. *Wall Street Journal*. <https://www.wsj.com/articles/the-end-of-typing-the-internets-next-billion-users-will-use-video-and-voice-1502116070>
- Benkler, Y. (2000). From Consumers to Users: Shifting the Deeper Structures of Regulation Toward Sustainable Commons and User Access. *Federal Communications Law Journal*, 52(3), 561–579.
- Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.
- Benkler, Y. (2013). Foreword. In M. L. Smith & K. M. A. Reilly (Eds.), *Open Development: Networked Innovations in International Development* (pp. vii–xiii). International Development Research Centre.
- Benkler, Y. (2017a). Peer production, the commons, and the future of the firm. *Strategic Organization*, 15(2), 264–274. <https://doi.org/10.1177/1476127016652606>
- Benkler, Y. (2017b). Law, Innovation, and Collaboration in Networked Economy and Society. *Annual Review of Law and Social Science*, 13(1), 231–250. <https://doi.org/10.1146/annurev-lawsocsci-110316-113340>
- Benkler, Y. (2017c). Peer Production and Cooperation. In J. M. Bauer & M. Latzer (Eds.), *Handbook on the Economics of the Internet* (pp. 91–119). Edward Elgar Pub.
- Benkler, Y., Shaw, A., & Hill, B. M. (2015). Peer Production: A Form of Collective Intelligence. In T. W. Malone & M. S. Bernstein (Eds.), *Handbook of Collective Intelligence* (1st ed., pp. 1–33). The MIT Press.
- Bentley, C. M. (2017). An Analysis of Accountability Concepts for Open Development. In J. Choudrie, M. S. Islam, F. Wahid, J. M. Bass, & J. E. Priyatma (Eds.), *Information and Communication Technologies for Development* (Vol. 504, pp. 793–802). Springer International Publishing. https://doi.org/10.1007/978-3-319-59111-7_64
- Bentley, C. M., & Chib, A. (2016). The Impact of Open Development Initiatives in Lower- and Middle Income Countries: A Review of the Literature. *The Electronic Journal of Information Systems in Developing Countries*, 74(1), 1–20. <https://doi.org/10.1002/j.1681-4835.2016.tb00540.x>
- Bentley, C. M., Chib, A., & Poveda, S. (2019). A critical narrative approach to openness: The impact of open development on structural transformation. *Information Systems Journal*, 29(4), 787–810. <https://doi.org/10.1111/isj.12226>

- Berdou, E. (2017). Open Development in Poor Communities: Opportunities, Tensions, and Dilemmas. *Information Technologies & International Development*, 13, 18–32.
- Blatter, J. K. (2008). Case Study. In L. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (Online Edition, pp. 68–72). SAGE Publications, Inc. <https://doi.org/10.4135/9781412963909.n39>
- Booth, D., & Golooba-Mutebi, F. (2012). Developmental patrimonialism? The case of Rwanda. *African Affairs*, 111(444), 379–403. <https://doi.org/10.1093/afrad/ads026>
- boyd, danah, & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Braybrooke, K., & Jordan, T. (2017). Genealogy, Culture and Technomyth: Decolonizing Western Information Technologies, from Open Source to the Maker Movement. *Digital Culture & Society*, 3(1), 25–46. <https://doi.org/10.14361/dcs-2017-0103>
- BTI. (2018). *Rwanda Country Report*. Bertelsmann Transformation Index. <https://www.bti-project.org/de/berichte/laenderberichte/detail/itc/RWA/>
- Buskens, I. (2013). Open Development is a Freedom Song: Revealing Intent and Freeing Power. In M. L. Smith & K. M. A. Reilly (Eds.), *Open Development: Networked Innovations in International Development* (pp. 327–351). International Development Research Centre.
- Carmody, P. (2013). A knowledge economy or an information society in Africa? Thintegration and the mobile phone revolution. *Information Technology for Development*, 19(1), 24–39. <https://doi.org/10.1080/02681102.2012.719859>
- Castells, M. (2010). *The Rise of the Network Society* (2nd ed.). Wiley-Blackwell.
- Cohen, J. E. (2013). What Privacy is for. *Harvard Law Review*, 126(7), 1904–1933.
- Costanza-Chock, S. (2018, June 3). Design Justice: Towards an intersectional feminist framework for design theory and practice. *Design Research Society 2018*. <https://ssrn.com/abstract=3189696>
- Costanza-Chock, S. (2020). *Design Justice: Community-led Practices To Build The Worlds We Need*. The MIT Press.
- Couldry, N., & Mejias, U. (2019a). Making data colonialism liveable: How might data's social order be regulated? *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1411>
- Couldry, N., & Mejias, U. A. (2019b). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- Couldry, N., & Mejias, U. A. (2019c). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, 20(4), 336–349. <https://doi.org/10.1177/1527476418796632>
- Couldry, N., & Yu, J. (2018). Deconstructing datafication's brave new world. *New Media & Society*, 20(12), 4473–4491. <https://doi.org/10.1177/1461444818775968>
- Creative Commons. (n.d.). *CC0 1.0 Universal*. CC Creative Commons. Retrieved 1 February 2020, from <https://creativecommons.org/publicdomain/zero/1.0/>
- Crisafulli, P., & Redmond, A. (2012). *Rwanda, Inc: How a devastated nation became an economic model for the developing world*. Palgrave Macmillan.

- Dahir, A. L. (2018, November 27). *African languages are being left behind when it comes to voice recognition innovation*. Quartz Africa. <https://qz.com/africa/1475763/african-languages-are-lagging-behind-when-it-comes-to-voice-recognition-innovations/>
- De Beer, J., & Oguamanam, C. (2013). Open Minds: Lessons from Nigeria on Intellectual Property, Innovation, and Development. In M. L. Smith & K. M. A. Reilly (Eds.), *Open Development: Networked Innovations in International Development* (pp. 249–272). International Development Research Centre.
- Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring Data Justice: Conceptions, Applications and Directions. *Information, Communication & Society*, 22(7), 873–881. <https://doi.org/10.1080/1369118X.2019.1606268>
- Digital Umuganda. (n.d.). *Building Common Digital Infrastructure for Rwanda* [Digital Umuganda [Website]]. Retrieved 3 February 2020, from <https://digitalumuganda.com/>
- Digital Umuganda, & GIZ. (2019). *Cooperation Agreement between Digital Umuganda Ltd and the GIZ Open Resources Incubator (ORI)* [Unpublished]. Digital Umuganda Ltd; Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ).
- Economist. (2017a, May 6). Data is giving rise to a new economy. *The Economist*. <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy>
- Economist. (2017b, May 6). The world's most valuable resource is no longer oil, but data. *The Economist*. <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>
- Elder, L., Samarajiva, R., Gillwald, A., & Galperin, H. (2013). *Information lives of the poor: Fighting poverty with technology*. International Development Research Centre.
- Freedom House. (2018, November 1). *Rwanda Country Report*. Freedom on the Net 2018. <https://freedomhouse.org/report/freedom-net/2018/rwanda>
- Friederici, N., Donovan, K., & Jama, A. (2013). *Towards Transformation? ICT in Post-Conflict Rwanda*. infoDev / World Bank. <http://www.infodev.org/articles/towards-transformationict-post-conflict-rwanda>
- Gagliardone, I., & Golooba-Mutebi, F. (2016). The Evolution of the Internet in Ethiopia and Rwanda: Towards a “Developmental” Model? *Stability: International Journal of Security and Development*, 5(1). <https://doi.org/10.5334/sta.344>
- Gimpel, L., Brumund, D., Stojanovic, A., & Krewer, J. (2019a, May 22). *Team Voice: Helping AI speak your local language* [Unpublished] [Presentation at ‘GIZ Innovation Fund: Final Pitch Event’]. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Bonn.
- Gimpel, L., Brumund, D., Stojanovic, A., & Krewer, J. (2019b). *Open Resources Incubator (ORI): Mobilising the power of open resources to reach the sustainable development goals* [Unpublished concept note]. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ).
- Gitelman, L. (Ed.). (2013). *‘Raw data’ is an oxymoron*. The MIT Press.
- GIZ. (n.d.-a). *Digital solutions for sustainable development (DSSD)*. Deutsche Gesellschaft Für Internationale Zusammenarbeit (GIZ). Retrieved 9 February 2020, from <https://www.giz.de/en/worldwide/73176.html>

- GIZ. (n.d.-b). *GIZ Innovation Fund*. Deutsche Gesellschaft Für Internationale Zusammenarbeit (GIZ). Retrieved 18 January 2020, from <https://www.giz.de/expertise/html/60797.html>
- Graham, M. (Ed.). (2019). *Digital economies at global margins*. The MIT Press; International Development Research Centre.
- Graham, M., & Haarstad, H. (2011). Transparency and Development: Ethical Consumption through Web 2.0 and the Internet of Things. *Information Technologies & International Development*, 7(1), 1–18.
- Grand View Research. (2018, November). *Speech And Voice Recognition Market Worth \$31.82 Billion by 2025*. Grand View Research. <https://www.grandviewresearch.com/press-release/global-voice-recognition-industry>
- Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2). <https://firstmonday.org/article/view/3316/2764>
- Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean? *Information Development*, 34(1), 90–102. <https://doi.org/10.1177/0266666916678282>
- Heeks, R., & Shekhar, S. (2019). Datafication, development and marginalised urban communities: An applied data justice framework. *Information, Communication & Society*, 22(7), 992–1011. <https://doi.org/10.1080/1369118X.2019.1599039>
- Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, 34(1), 135–174. <https://doi.org/10.1111/dpr.12142>
- Irani, L. (2019). Hackathons and the Cultivation of Platform Dependence. In M. Graham (Ed.), *Digital economies at global margins* (pp. 223–248). The MIT Press; International Development Research Centre.
- ITU. (2018). *Measuring the Information Society (Volume 2)*. International Telecommunications Union. <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/misr2018.aspx>
- Jackson, S. J. (2014). Rethinking Repair. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 221–239). The MIT Press.
- Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, 32(5), 694–711. <https://doi.org/10.1177/0894439314525902>
- Just Net Coalition. (2019). *Digital Justice Manifesto: A Call to Own Our Digital Future*. Just Net Coalition. <https://justnetcoalition.org/digital-justice-manifesto.pdf>
- Kaliisa, S. P. (2019, May 20). *New digital transformation centre 'a relief to local tech innovators'*. The New Times | Rwanda. <https://www.newtimes.co.rw/business/new-digital-transformation-centre-relief-local-tech-innovators>
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE Publications.
- Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing. *Science, Technology & Innovation Studies*, 4(1), 5–26. <https://doi.org/10.17877/de290r-12790>

- Kleppe, A., & Gimpel, L. (2019, February 26). *Sustainable tech development needs local solutions: Voice tech ideation in Kigali*. Medium. <https://medium.com/mozilla-open-innovation/sustainable-tech-development-needs-local-solutions-voice-tech-ideation-in-kigali-ec8dd33e0823>
- Lanier, J. (2014). *Who owns the future?* (Publ. in Penguin Books with updates). Penguin Books.
- Linksvayer, M. (2013). Creative Commons: Governing the Intellectual Commons from Below. *The Wealth of the Commons*. <http://wealthofthecommons.org/essay/creative-commons-governing-intellectual-commons-below>
- Longman, T. (2011). Limitations to Political Reform: The Undemocratic Nature of Transition in Rwanda. In S. Straus & L. Waldorf (Eds.), *Remaking Rwanda: State building and human rights after mass violence* (pp. 25–47). The University of Wisconsin Press.
- Mann, L., & Berry, M. (2016). Understanding the Political Motivations that Shape Rwanda's Emergent Developmental State. *New Political Economy*, 21(1), 119–144.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. John Murray Publishers.
- Mazzucato, M. (2018a). *The Value of Everything: Making and Taking in the Global Economy* (01 ed.). Allen Lane.
- Mazzucato, M. (2018b, December 21). *Mariana Mazzucato on who creates value* [FT Alphachat]. <https://www.ft.com/content/219b82d5-5041-4529-87b5-a6edc1d35c81>
- Mbembe, A. (2017, January 6). *The digital age erases the divide between humans and objects*. Mail & Guardian. <https://mg.co.za/article/2017-01-06-00-the-digital-age-erases-the-divide-between-humans-and-objects/>
- McGee, R., & Pettit, J. (Eds.). (2019). *Power, Empowerment and Social Change*. Routledge.
- McKernan, L. E. F. (2008). Participant observation. In L. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412963909.n249>
- Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1428>
- Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television & New Media*, 20(4), 319–335. <https://doi.org/10.1177/1527476419837739>
- Mizukami, P. N., & Lemos, R. (2008). From Free Software to Free Culture: The Emergence of Open Business. In L. Shaver (Ed.), *Access to Knowledge in Brazil: New Research on Intellectual Property, Innovation and Development* (pp. 13–40). Bloomsbury Academic.
- Morozov, E. (2018a, March 31). After the Facebook scandal it's time to base the digital economy on public v private ownership of data. *The Observer*. <https://www.theguardian.com/technology/2018/mar/31/big-data-lie-exposed-simply-blaming-facebook-wont-fix-reclaim-private-information>

- Morozov, E. (2018b, August 19). There is a leftwing way to challenge big tech for our data. Here it is. *The Guardian*. <https://www.theguardian.com/commentisfree/2018/aug/19/there-is-a-leftwing-way-to-challenge-big-data-here-it-is>
- Morozov, E. (2019, February 27). The left needs to get radical on big tech – moderate solutions won't cut it | Evgeny Morozov. *The Guardian*. <https://www.theguardian.com/commentisfree/2019/feb/27/left-radical-big-tech-moderate-solutions>
- Morris, N. (2005). The diffusion and participatory models: A comparative analysis. In O. Hemer & T. Tufte (Eds.), *Media & Glocal Change: Rethinking Communication for Development* (pp. 123–144). CLACSO Books. <http://bibliotecavirtual.clacso.org.ar/ar/libros/edicion/media/12Chapter7.pdf>
- Mozilla. (n.d.-a). *Common Voice: Datasets*. Common Voice [Website]. Retrieved 4 February 2020, from <https://voice.mozilla.org/en/datasets>
- Mozilla. (n.d.-b). *Common Voice: Homepage*. Common Voice [Website]. Retrieved 29 January 2020, from <https://voice.mozilla.org/>
- Mozilla. (2019, November 25). *Mozilla and BMZ Announce Cooperation to Open Up Voice Technology for African Languages*. The Mozilla Blog. <https://blog.mozilla.org/blog/2019/11/25/mozilla-and-bmz-announce-cooperation-to-open-up-voice-technology-for-african-languages>
- Murphy, J. T., & Carmody, P. R. (2015). *Africa's information revolution: Technical regimes and production networks in South Africa and Tanzania*. Wiley/ Blackwell.
- MYICT. (2015). *Smart Rwanda 2020 Master Plan: Towards a Knowledge-based Society*. Ministry of Youth and ICT. https://minict.gov.rw/fileadmin/Documents/Mitec2018/Policies___Publication/Strategy/SMART_RWANDA_MASTER_PLAN_FINAL.pdf
- MYICT. (2017). *National Data Revolution Policy*. Ministry of Youth and ICT.
- Nibeza, S. (2015). ICT4D, A Driver of Socio-Economic Development in Rwanda: Situation 2000-2015. *Research Journal of Economics & Business Studies*, 5(1), 19–34.
- Nickolls, B. (2017). *A One Day Conversation for Open Source Software Sustainers*. Sustain Conference, San Francisco. <https://sustainoss.org/assets/pdf/SustainOSS-west-2017-report.pdf>
- NISR. (2018). *Integrated Household Living Conditions Survey (EIVC 5) 2016-2017*. National Institute of Statistics Rwanda. <https://microdata.statistics.gov.rw/index.php/catalog/82>
- Ntale, A., Yamanaka, A., & Nkurikiyimfura, D. (2013). The Metamorphosis to a Knowledge-Based Society: Rwanda. In B. Bilbao-Osorio, S. Dutta, & B. Lanvin (Eds.), *The Global Information Technology Report 2013* (pp. 119–125). World Economic Forum.
- Nussbaum, M. C. (2013). *Creating Capabilities: The Human Development Approach* (Reprint). Harvard University Press.
- Nyirishema, P. (2017, October 10). *Policy Statement*. World Telecommunication Development Conference 2017. https://www.itu.int/en/ITU-D/Conferences/WTDC/WTDC17/Pages/High-Level-Segment-Speeches.aspx?SpeechID=164&View=%7B935647de-b11b-4fc5-9fbe-9f36ef9cf04a%7D&SortField=Session_x003a_SessionNameEN&SortDir=Asc

- O'Connor, G. C., Corbett, A. C., & Peters, L. S. (2018). *Beyond the Champion: Institutionalizing Innovation through People*. Stanford Business Books, an imprint of Stanford University Press.
- OECD. (2015). *Data-Driven Innovation: Big Data for Growth and Well-Being*. OECD. <https://doi.org/10.1787/9789264229358-en>
- Ojanperä, S., Graham, M., & Zook, M. (2019). The Digital Knowledge Economy Index: Mapping Content Production. *The Journal of Development Studies*, 1–18. <https://doi.org/10.1080/00220388.2018.1554208>
- Pieterse, J. N. (2010). *Development Theory: Deconstructions/Reconstructions* (2nd ed). SAGE.
- Plotkin, D. (2014). *Data stewardship: An actionable guide to effective data management and data governance*. Elsevier/Morgan Kaufman.
- Pomerantz, J., & Peek, R. (2016). Fifty shades of open. *First Monday*, 21(5). <https://doi.org/10.5210/fm.v21i5.6360>
- Purdeková, A. (2016). “Mundane Sights” of Power: The History of Social Monitoring and Its Subversion in Rwanda. *African Studies Review*, 59(2), 59–86. <https://doi.org/10.1017/asr.2016.32>
- Reilly, K. M. A., & Alperin, J. P. (2016). Intermediation in Open Development: A Knowledge Stewardship Approach. *Global Media Journal*, 9(1), 51–71.
- Reilly, K. M. A., & Smith, M. L. (2013). The Emergence of Open Development in a Network Society. In M. L. Smith & K. M. A. Reilly (Eds.), *Open Development: Networked Innovations in International Development* (pp. 15–50). International Development Research Centre.
- Rubin, H., & Rubin, I. (2005). *Qualitative Interviewing: The Art of Hearing Data* (2nd edition). SAGE Publications, Inc. <https://doi.org/10.4135/9781452226651>
- Samuelson, B. L., & Freedman, S. W. (2010). Language policy, multilingual education, and power in Rwanda. *Language Policy*, 9(3), 191–215. <https://doi.org/10.1007/s10993-010-9170-7>
- Scholz, L. (2018). *Big Data is Not Big Oil: The Role of Analogy in the Law of New Technologies* [Research Paper]. Florida State University College of Law.
- Schrape, J.-F. (2019). *Technology and the Promise of Decentralization. Origins, Development, Patterns of Arguments* (SOI Discussion Paper No. 2019–01; Stuttgarter Beiträge Zur Organisations- Und Innovationsforschung). Universität Stuttgart, Institut für Sozialwissenschaften, Organisations- und Innovationssoziologie. <http://hdl.handle.net/10419/194289>
- Sebarenzi, J., & Twagiramungu, N. (2019, April 8). *Rwanda's economic growth could be derailed by its autocratic regime*. The Conversation. <http://theconversation.com/rwandas-economic-growth-could-be-derailed-by-its-autocratic-regime-114649>
- Sen, A. (2001). *Development as Freedom* (New Ed). Oxford University Press.
- Singh, P. J., & Gurumurthy, A. (2013). Establishing Public-ness in the Network: New Moorings for Development—A Critique of the Concepts of Openness and Open Development. In M. L. Smith & K. M. A. Reilly (Eds.), *Open Development: Networked Innovations in International Development* (pp. 173–196). International Development Research Centre.

- Smith, M. L. (2015, May). Being Open in ICT4D. *SSRN Electronic Journal*. 13th International Conference on Social Implications of Computers in Developing Countries, Negombo, Sri Lanka. <https://doi.org/10.2139/ssrn.2526515>
- Smith, M. L., Elder, L., & Emdon, H. (2011). Open Development: A New Theory for ICT4D. *Information Technologies & International Development*, 7(1), iii–ix.
- Smith, M. L., & Reilly, K. M. A. (Eds.). (2013). *Open Development: Networked Innovations in International Development*. International Development Research Centre.
- Smith, M. L., & Seward, R. (2017). Openness as social praxis. *First Monday*, 22(4). <https://doi.org/10.5210/fm.v22i4.7073>
- Takeuchi, S. (2019). Development and Developmentalism in Post-genocide Rwanda. In Y. Takagi, V. Kanchoochat, & T. Sonobe (Eds.), *Developmental State Building* (pp. 121–134). Springer Singapore. https://doi.org/10.1007/978-981-13-2904-3_6
- Taylor, L., & Broeders, D. (2015). In the name of Development: Power, profit and the datafication of the global South. *Geoforum*, 64, 229–237. <https://doi.org/10.1016/j.geoforum.2015.07.002>
- Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, 34(6), 990–1006. <https://doi.org/10.1177/0263775816633195>
- Thomson, S. (2018). *Rwanda: From Genocide to Precarious Peace*. Yale University Press.
- Tufte, T. (2017). *Communication and Social Change: A Citizen Perspective*. Polity.
- UN Global Pulse. (2012). *Big Data for Development: Challenges & Opportunities*. United Nations. <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>
- Unwin, T. (2017). *Reclaiming Information and Communication Technologies for Development*. Oxford University Press.
- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208. <https://doi.org/10.24908/ss.v12i2.4776>
- van Schalkwyk, F., Canares, M., Chattapadhyay, S., & Andrason, A. (2015). *Open Data Intermediaries in Developing Countries*. World Wide Web Foundation. <https://webfoundation.org/research/open-data-intermediaries-in-developing-countries/>
- van Schalkwyk, F., Willmers, M., & McNaughton, M. (2016). Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development*, 22(sup1), 68–83. <https://doi.org/10.1080/02681102.2015.1081868>
- von Hippel, E. (2005). *Democratizing Innovation*. The MIT Press.
- Wilkins, K. G. (2015). Development Communication. In W. Donsbach (Ed.), *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781405186407.wbiecd020.pub2>
- World Bank. (2018). *Information and Communications for Development 2018: Data-Driven Development*. The World Bank. <https://doi.org/10.1596/978-1-4648-1325-2>
- World Bank. (2019). *Rwanda: Systematic Country Diagnostic*. World Bank. <http://documents.worldbank.org/curated/en/219651563298568286/pdf/Rwanda-Systematic-Country-Diagnostic.pdf>

- World Economic Forum. (2012). *Big Data, Big Impact: New Possibilities for International Development*. World Economic Forum. <https://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>
- Zuboff, S. (2019, July 2). It's not that we've failed to rein in Facebook and Google. We've not even tried. *The Guardian*. <https://www.theguardian.com/commentisfree/2019/jul/02/facebook-google-data-change-our-behaviour-democracy>
- Zukin, S., & Papadantonakis, M. (2017). Hackathons as Co-optation Ritual: Socializing Workers and Institutionalizing Innovation in the "New" Economy. In A. L. Kalleberg & S. P. Vallas (Eds.), *Research in the Sociology of Work* (Vol. 31, pp. 157–181). Emerald Publishing Limited. <https://doi.org/10.1108/S0277-283320170000031005>