



A Prototype Data Governance Framework for Africa

Bitange Ndemo and Aaron Thegeya

2.1 INTRODUCTION

In their simplest form, data are frequently defined as a collection of symbols that are the properties of observables or the representation of facts. Data within a given context translate into information—and information in perspective, integrated into a viewpoint based on experience—is what we think of as knowledge (Ackoff, 1989). Despite the distinction between data and information, the terms are often interchangeable in practice. Data are an important component of total factor productivity and contribute in important ways to growth, in addition to labor and capital. There are massive economies of scale to be gained from combining different data sets to yield insights that would be otherwise unavailable or difficult to capture. In addition, improvements in data processing, data storage, and data analytics through machine learning and artificial intelligence can support

B. Ndemo (✉)

Kenya's Ambassador, Belgium and the EU Mission, Brussels, Belgium
e-mail: bndemo@bitangendemo.me

A. Thegeya

Aliquot Limited, Nairobi, Kenya

© The Author(s) 2023

B. Ndemo et al. (eds.), *Data Governance and Policy in Africa*,
Information Technology and Global Governance,
https://doi.org/10.1007/978-3-031-24498-8_2

productivity gains, boost efficiency, and decrease costs—advances that can drive economic growth, increase prosperity, and improve the standard of living on the continent.

Data governance involves establishing principles to enable an environment for the sharing of data, with the ultimate goal of improving living standards, while at the same time recognizing and protecting the rights of data originators and users. Given the central role of data in today's global economy, a system of effective data governance is essential. That said, the development of any such framework requires careful scrutiny of the economic, legal, and institutional issues attendant to such regulation, as well as the establishment of proper standards for the exchange and protection of data.

At the micro or firm level, data governance has historically referred to managing the availability, usability, integrity, and security of data. From a global perspective, the World Bank (2021a) has deemed that data governance “entails creating an environment of implementing norms, infrastructure policies and technical mechanisms, laws and regulations for data, related economic policies, and institutions that can effectively enable the safe, trustworthy use of data to achieve development outcomes” (p. 38). To leverage the vast opportunities of data utilization, Africa must develop a data strategy underpinned by a governance framework. Such a strategy would establish data sovereignty and render the continent more competitive and better positioned to engage in cross-country collaboration during the digital age. Data could be reused by promoting practices protective of privacy, including personal and other sensitive data, through techniques including anonymization, pseudonymization, differential privacy, generalization, suppression, and randomization.

Africa must define a continental strategy and devise a governance framework that maximizes the use of data while ensuring productive cross-border data flows and protecting individual rights. The continent lacks sturdy and expansive national or regional structures for governing data, and individual African nations have yet to develop legislation to safeguard data use and digital transactions—an absence likely to cause market fragmentation due to insufficient harmonization (United Nations Congress on Trade and Development, 2021). Meanwhile, the data governance frameworks that do exist, albeit in their limited form, lack coherence in terms of principles, scope, and enforceability across jurisdictions. The rapidly changing landscape of data generation, storage, and mining capacity—as well as the dearth of human and financial resources, reliable institutions, and

enforcement capacity to support an efficient data governance environment—will, absent immediate action by key stakeholders, cause the continent to regress at the moment when it is arguably positioned to show its greatest progress ever.

Without a coherent data governance framework, data generated in Africa risk being improperly utilized within each country and in other parts of the world, leading to an unbalanced platform of data exchange with countries where data are closely regulated. African countries, therefore, must take the lead in establishing the appropriate frameworks that will serve not only their own national interests but also those of the continent as a whole. Governments, development institutions, and nonstate actors should collaborate to implement and enforce data governance laws and policies that can make the continent’s digital economy more competitive while, at the same time, enhancing transparency, trust, and digital inclusiveness for all users.

2.2 BACKGROUND AND LITERATURE REVIEW

The history of data is closely intertwined with the evolution of mankind. The earliest examples of data being stored and analyzed by humans date back to about 18,000 BCE, in what is now Uganda, when humans were recorded using the Ishango bone for the purposes of tallying (Marr, 2015). The bones were marked with notches to keep track of trading activity, and notches were compared between bones to carry out rudimentary calculations on supplies. Subsequently, the abacus, the first device constructed specifically for performing calculations, was invented around 2400 BCE. The first data libraries appeared during roughly the same period, marking mankind’s initial endeavor toward mass data storage. The year 1663 saw the emergence of statistics as a distinct mode of analysis, when John Graunt recorded mortality information in London and used his figures and framework to design an early warning system to alert the population about the spread of the bubonic plague that had been ravaging Europe. The central concept of the modern computer emerged thereafter, based on the ideas of Alan Turing, who, in 1936, presented the notion of a universal machine (Zimmermann, 2017), paving the way for the first digital computers in the following decade. Finally, data became ubiquitous with the advent of the Internet, announced by Tim Berners-Lee, in 1991, thus setting the stage for the modern age of big data.

Historically, people struggled to collect data because they lacked the necessary tools and infrastructure; the digital revolution, however, led to dramatic changes in the scope and types of data collected, and the volume of data sets collected has increased compared to only a few decades ago. What's more, when governments fail to do the collecting, private firms and individuals can now use new digital platforms to gather data for private use, for commercial purposes, or to promote accountability and governance—such as platforms used to report violence or discrimination. The cataloguing of information from Africa's past, through the digitization of archived records and the utilization of disparate data sources for analysis of economic activity, climate, and terrain, for example, has increased the set of data available as well as analytical findings from both research and commercial perspectives. This transformation has encouraged insightful publications on Africa's past (see, e.g., Fourie, 2016) that would not have been possible without today's data infrastructure.

Data consumption needs have increased significantly over time, and consumption of data varies by region. Daily usage statistics are staggering: From the advent of civilization to 2003, for example, five exabytes of data were created¹; but, by only seven years later, that amount of data was being generated every two days (World Bank, 2021b). By 2025, it is estimated that 463 exabytes of data will be created around the world each day (Desjardins, 2019). Presently, the entire universe of data is estimated at 44 zettabytes, a total that accounts, for example, for the 294 billion emails sent, the 5 billion Internet searches that occur, and the 65 billion messages transmitted each day through messaging services such as WhatsApp (World Bank, 2021b).

A World Bank (2021b) study looking at minimum data consumption using data from six developing and emerging countries found that the most frequent online activities, which included visits to public service websites, learning, shopping, health information and news, consumed 660 megabytes of data per user, per month. When looking beyond data requirements for solely welfare-improving activities like those just mentioned, individuals in these countries needed an additional 5.2 gigabytes for recreational activities on social media per month, putting total monthly data demand in these economies at approximately 6 gigabytes per person (2021b). Never has a data governance framework been more necessary than it is now.

¹An exabyte corresponds to 10^{21} bytes, and a zettabyte corresponds to 10^{24} bytes.

A review of the literature shows that the body of research on data governance has been carried out mostly from an organizational perspective. Given data's role as a strategic and monetizable asset, organizations have researched holistic data governance frameworks to facilitate effective utilization of data with a profit motive, while respecting privacy rights (see, e.g., Khatri & Brown, 2010; Otto, 2011; Weber et al., 2009). From a regulatory perspective, countries are in the process of defining data governance frameworks. For example, in November 2020, the European Commission proposed rules on data governance to boost data sharing and support European data spaces, in line with principles such as personal data protection (General Data Protection Regulation), consumer protection, and competition. The World Bank has even focused its 2021 World Development Report on data issues pertinent to developing economies.

Micheli et al. (2020) investigated the emerging models of data governance in the age of datafication and, in addressing the politics of data, considered actors' competitive struggles. This conceptualization brought to the forefront the multifaceted economic and social interactions, as well as power relations, within data governance models—particularly those at work in corporate environments. Public bodies and civil society are, within these models, key players for both redistributing any value produced via data and democratizing its governance. Further, Micheli et al. found that data trust and intermediaries were included in nearly every investigated model, leading the researchers to underscore the importance of data infrastructure as fundamental to improving trust in data.

Research has also revealed a wide variety of views and minimal agreement across stakeholders on the issue of data governance frameworks. Within the context of academia, Kouper et al. (2020) carried out an exploratory study on data governance in the United States, involving individuals who worked in research and academic institutions, aiming to understand the entities central to decision-making and governance on data and research-related issues. This group's findings showed considerable complexity and diversity across stakeholders in terms of both identity and ideas on the governance of data. To account for such diversity, Kouper et al. proposed to frame data governance in research around common governance bodies, arguing as well that, to ensure effective data governance in research, voices of people from different literacy and income levels should always be incorporated in shaping policy and making decisions.

Several approaches have been used to determine data governance activities. For instance, Alhassan et al. (2016) used key words to identify papers

on data governance activities using open-coding approaches and identified 31 articles that mentioned such activities. Their analysis identified 110 data governance activities across five decision domains of their framework (data principles, metadata, data quality, data life cycle, and data access), with each domain implicating a different critical aspect of data governance.

The rapid growth in digital financial services presents concerns over data protection and privacy for low-income individuals, especially those in developing countries. Vidal and Medine (2019), for example, analyzed whether data privacy is desirable in a corporate world. Their analysis included experiments in India and Kenya, where several products with varying degrees of data protection and a range of privacy options were offered to low-income individuals, thereby allowing the researchers to evaluate the demand for individual safeguards within markets with limited or no frameworks in place to protect individual privacy, and they found that low-income individuals were willing to pay for their data privacy. For instance, in Kenya, 64% of low-income individuals surveyed chose options with a greater degree of data privacy, despite the imposition of a non-trivial 10% fee attached to this option. Even more, results in Bangalore were similar to those of Kenya, with 66% of survey participants choosing this option.

Freely available public data could generate economies of scale through reuse, and the benefits of these types of data in terms of the public good present a case for protecting the availability of some classes of data from public sources relative to private firms. Beraja et al. (2020) analyzed the state of artificial intelligence within China by gathering comprehensive data from government and firm-procurement contracts within the artificial intelligence industry and found that sharing data improved productivity in both private and public institutions. Their results also indicated that the ability to access government data outweighed the feasibility of providing these same data through commercial means; accessible government data should not, they concluded, be substituted by private markets.

2.3 AN ORGANIZATIONAL FRAMEWORK FOR DATA GOVERNANCE

A sound data governance framework requires that institutions and stakeholders have the right incentives to produce, protect, and share data; a comprehensive understanding of data governance also demands

consideration of key dimensions including (a) the relevant stakeholders who use data and those who are impacted by the use of data; (b) the life cycle of data from creation to destruction; (c) the typology of data, reflecting relevant characteristics that impact processing, storage, and accuracy; and (d) enabling pillars such as economic, legal, and institutional aspects that create the necessary infrastructure for using data and maximize its productivity. These key dimensions are illustrated in Fig. 2.1.

The key stakeholders who generate and use data include households, the private sector, governments, and civil society, with households and the private sector being major data producers and/or consumers and

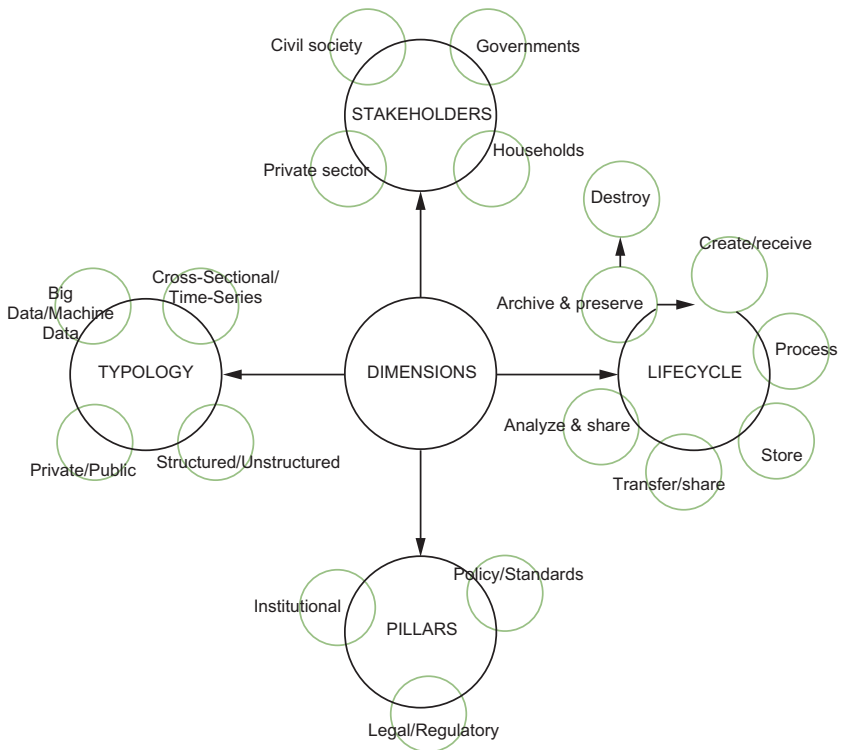


Fig. 2.1 Organizational framework for data governance. Note: Figure conceived and designed by the authors

governments and civil society offering essential safeguards concerning its use. The government is central in formulating policies, and regulations, while civil society helps hold other stakeholders accountable. The needs of each stakeholder bear consideration within a data governance framework. Data privacy concerns, for example, vary across stakeholders and are relevant particularly for households and the private sector; conversely, certain data collected by governments merit classification under public data, particularly where the utilization of these data improves productivity and creates economies of scale—and also given that the data are collected using public resources.

The data life cycle details the key steps that occur between the creation and destruction or reuse of data, specifically the collection, processing, and storage of data; transferring or sharing of data among users; analysis and value addition; archiving and preservation for future use; and destruction of data at the end of the cycle. Stored or archived data are usually available for reuse, and an enabling infrastructure is essential during each step, including security of storage and transmission through encryption—protocols that enable data transfer across systems, allow its destruction at the end of the cycle, and maintain integrity and accuracy of data by preventing unauthorized manipulation.

Data collection and processing methods help determine accuracy, in turn promoting greater trust in data sets. Established data collection techniques for public data include the collection of population statistics by an official authority, or the collection of sample statistics using rigorous sample-design techniques. These methods yield accurate and trusted data sets structured in nature, but they also tend to command significant resources during the collection, depending on the level of disaggregation required, either in terms of subpopulations or regions of interest. Due to the financial cost, as well as planning and logistical requirements associated with collection, these techniques tend to be implemented infrequently. Therefore, analyses based on these data usually have gaps, either in their level of disaggregation or across time.

A large number of distinct typologies of data exist, determined by the multidimensional aspects inherent in data, as well as the lens or perspective through which data are viewed. Data can be classified according to whether they are for private or public use, a distinction that, in turn, determines how widely available they might be as well as their cost to access. Data collected for commercial use are treated as a private good, and those who own them enjoy a competitive advantage as well as the ability to collect

fees when selling them. Public data, by contrast, which were collected using public resources, are intended to be widely available. Publicly available data usually provide social value and are useful inputs for other economic activities. Open data, for example, are a type of public data shared to fortify public governance and increase transparency, while also generating commercial opportunities.

For purposes of classification, structured data are organized according to some predefined model and stored electronically, typically in a relational database within a tabular format. Databases allow for efficient searching, editing, and error detection; they can also be more easily manipulated by programming languages. Conversely, unstructured data are less organized, are typically text-heavy, and require more flexible data structures. These data are more difficult for programs to process. Additionally, data can be classified according to cross-sectional and temporal dimensions, with cross-sectional data including many observations on subjects recorded at a fixed point and time-stamped data accounting for observations on one or many subjects recorded over time. Spatio-temporal data describe both the time and location of a particular event.

The utilization of big data is a subset of new collection and analysis techniques involving unstructured data, techniques made possible by accessible, less expensive, and more expansive storage capacity, as well as by advances in machine learning related to processing capacity and the increasing production of large amounts of digital data. These techniques reveal patterns from high frequency data, in real time, while low statistical errors within the data are supported by the large number of observations. Machine-learning algorithms depend on the availability of large data sets, with the predictive power of the algorithms increasing as the data become more available, even as the effectiveness of the algorithms continues to depend on the accuracy of the training data being used.

Newer techniques for collecting data rely on the availability of digital data and depend on both advances in machine learning and estimation theory at the small-area level. These methods offer advantages relative to traditional methods in terms of cost, frequency, and coverage; the use of small-area estimation techniques, for example, allows interpolation of statistics at a disaggregated level based on the combination of population, sample, and even satellite data. These methods may also include data from unstructured sources, and the accuracy of these methods remains an area of active research.

Underpinning the rights of stakeholders, the flow of data within its life cycle and the various data typologies are the enabling pillars of an effective data governance framework. The pillars include the economic, legal, and institutional framework that facilitates policies enabling the appropriate use of data while protecting data privacy, standards that embed data accuracy and make possible the secure storage and transfer of data, and the implementation and enforcement of appropriate regulation for the use of data. Establishing appropriate legally empowered institutions to create and regulate the data space is a critical dimension of the enabling framework.

A number of data governance frameworks exist, varying in terms of membership and degree of implementation. In 2014, the African Union adopted the Malabo Convention, which sought to encourage cybersecurity and personal data protection among partner countries, although the Convention was not fully implemented and thus not enforceable. According to the Convention, data need not be stored once the purpose for which it had been collected was met. This would have called for personal data to be protected by deletion when its purposes were achieved, meaning that data controllers would need to follow up with other data users to ascertain the destruction of personal data.

The Economic Community of West African States (ECOWAS) was established to promote the integration and economic growth of its member states. The member states adopted a Personal Data Protection Act in 2010, an agreement covering personal data and consent by the subject, recipient, and third parties, as well as the role of data processors and a data protection authority. However, the Act does not cover other important dimensions, such as profiling, anonymization, personal data breaches, and pseudonymization—matters of particular relevance with respect to cross-border data flows within the ECOWAS community. The Act requires member states to develop independent data-processing agreements for their citizens, guaranteeing their professional secrecy, impartiality, and power to punish errant parties. According to the Act, the processing of personal data is legitimate when carried out with the owner's consent and approval.

The Asia-Pacific Economic Cooperation (APEC) has developed a privacy framework for Asian countries, specifically in the Pacific region. APEC aims to promote flexible and effective information flows within the APEC community, while ensuring well-managed data protection. In 2020, to protect their government institutions, firms, and individuals against harm

or the risk of private data being exposed, as well as to promote trade and ensure trust among member states, New Zealand, Chile, and Singapore signed a Digital Economy Partnership Agreement (DEPA) governing and protecting the sharing and processing of electronic data.

Other well-established data governance frameworks can be found in the European Union and the United States. In 2018, the European Union put in place its General Data Protection Regulation (GDPR), enshrining it as the legally recognized framework for data privacy and protection among member states. This framework governs the European Union's member states and their trading partners in all matters of data governance. In a rather different and definitively disaggregated manner, the United States offers both state and federal laws to protect personal online data and privacy.

2.4 A PROTOTYPE DATA GOVERNANCE FRAMEWORK FOR AFRICA

Establishing an effective data governance framework for Africa requires a clear delineation of its objectives and careful attention to the unique characteristics of the continent. Africa has a large informal sector, an agricultural sector that dominates in production, and most of its commercial entities are small businesses. Much of the population connects through mobile phones, even as data access levels are much lower, averaging about 20% of the population, and while access to high speed data connections is even lower, still. Additionally, data access across households is highly uneven and depends on geographic location and economic status. For most Africans, the costs of enjoying Internet access are prohibitive, meaning that uneven access to data at a national level is mirrored by a large disparity in access at the continental level.

Digital-format data are highly limited in Africa; many public data sets are not digitized, and wide access to those that are digitized is low, fragmented, and inconsistent. Low levels of Internet connectivity also deter households and the private sector from generating new digital data, which, in turn, presents a major barrier to producing the high levels of data concentration that can spur innovative activity, enable the utilization of big data and machine learning techniques for data mining, and increase productivity. Data strategies and governance frameworks do not exist in many African countries, and, where frameworks do exist, they are typically

incomplete, disjointed, or not fully aligned with other existing and relevant legislation already in place, such as laws protecting individual rights. What's more, due to weak institutions, governance issues, or limited capacity, levels of enforceability within existing frameworks are low across the continent. The countries that already have data governance frameworks in place require close levels of coordination to avoid suffering fragmentation and, as a consequence, diminished effectiveness.

A pan-African data governance infrastructure can help the continent realize a single market for data, thereby enabling the creation, use, and reuse of data by individuals across Africa and spurring economic growth and development while protecting the rights of data subjects. A necessary prerequisite for a single data market is the generation of sufficient data to allow economies of scale through utilization. This means that an appropriate framework to rapidly increase data digitization and widespread access must be developed in parallel with a data governance framework, in addition to the establishment of other key legal and regulatory frameworks that comprehensively govern the data life cycle. Realizing an effective data governance framework is contingent on the establishment of country-level guidelines that provide a template instructing nations on the precise components necessary for a comprehensive framework, while also establishing principles to ensure coherence across the components within a country. Further, a complementary overall framework linked to and interoperable with national frameworks should be established at the continental level.

An effective framework requires a clear set of definitions and categories for different types of data as well as rules pertaining to the use and reuse of data within each category. In this regard, a framework should clearly define private versus public data and should offer clear guidelines on the use of each type. But effective implementation should also go a step further, designating key public data sets to be shared both nationally and across borders—data sets that should be identified according to the strategic interests of countries, thereby calling for a concurrent effort to determine and prioritize interests that will maximally promote the sharing of data. These may include, for example, expanding regional trade, boosting agricultural productivity and promoting food security, or dealing with climate-related threats. Cross-border data sharing can leverage principles employed in existing systems that effectively utilize information transcending borders, such as monitoring systems for infectious diseases.

A comprehensive data governance framework must rest on the widespread engagement of all stakeholders in a social contract that defines the protection of individual data, thereby building trust, creating an enabling environment that adds value to data, and promoting an equitable system (World Bank, 2021a). Such a social contract could overcome negative externalities resulting in the underutilization of data for productive activity, and, if properly implemented, it could define the role of and cultivate trust in data intermediaries—those figures or institutions central to the eventual success of a data governance framework.

Civil society has a central role to play in shaping the social contract across all other stakeholders by influencing policies on appropriate and optimal levels of data openness, transparency, and accountability. Through civil technology innovations such as open government platforms, civil society can leverage big data to improve tracking of the performance of public institutions in fulfilling their mandates, ensure wider dissemination of performance data to citizens to improve transparency, and integrate community feedback to allow broader active citizen participation in local government. Additionally, as the level of digitalization increases, civil society has a central role to play in ensuring that all individual rights are protected, in particular the rights of vulnerable individuals or those without an adequate level of awareness about their individual rights. Additionally, civil society is key to ensuring that big data is leveraged to increase the base of opportunities available to all individuals, and equally that digitalization does not increase income and data inequality.

Key elements of data property rights include guidance on the establishment of data ownership, as well as the appropriate level of control on data sharing. Property rights management is a critical part of any data management process; data owners have an interest in understanding how other users will utilize their data, and they also seek to ensure that ethical, legal, and professional obligations are observed. Data property rights are also important from the perspective of equity, as poor legal and governance structures can encourage misuse of information and render vulnerable those who enjoy neither authority nor influence.

A data governance framework can take a number of perspectives on data ownership, either creating a centralized authority responsible for monitoring and enforcing data-sharing regulations or following a more decentralized framework where data sharing resides at the individual level. Within this context, individual preferences can be brought to bear in terms

of the utility each individual obtains from maintaining data privacy, relative to the advantages that may accrue from data sharing, such as better matching and personalization of services. The appropriate framework can be implemented by passing on the control of data access protocols to individual users—for example, by enabling individuals to choose their level of access to different types of information generated by their devices. Indeed, privacy can be fragile and fleeting when third parties have access to sensitive data. When users share their data on different online platforms, they reveal signals about other users' preferences, based on shared exogenous characteristics. For example, the preferences of a teenager of a given age in a given school may signal the preferences of that person's circle of friends, perhaps creating negative externalities against those holding information about their preferences shared without their approval and limiting the scope of their control over personal information (Acemoglu et al., 2019).

Market failure may arise due to a lack of data rights. Data are non-rival and excludable,² creating incentives to hoard data and allowing the collection of rents as well as the maintenance of a dominant market position. In such situations, significant positive externalities to data sharing that could have a major impact on economic growth may fail to occur. In addition, organizations that collect data lack sufficient incentives to protect the privacy of users who have shared data, given that they do not internalize users' utility from privacy. In such cases, oversharing of data may occur.

With the increasing digitalization of information and improvements in algorithmic analysis, large amounts of data have been collected by companies in order to keep track of individual behavioral patterns, thereby enabling profiling and prediction of future actions. These data are then sold to third-parties for the purposes of monetization, by allowing these entities to market products more effectively to individuals. Enabled by a weak regulatory environment, individual data have been collected from devices and transmitted to companies without the awareness or consent of individuals, and thereafter sold to third-parties. This has prompted responses from various regulatory authorities, with some emphasizing the protection of individual privacy, and others leveraging the power of analytics and data to implement innovations such as social credit systems, which effectively increase surveillance over individuals. However, effective

²Non-rivalry of data means that data can be consumed or processed by multiple users without depleting its quality and/or supply. Excludability of data occurs when some groups or individuals are excluded from accessing or using the data.

regulation is complicated by the likelihood that regulatory authorities are behind the curve of innovative activity within major technology firms. Additionally, major technology corporations exert significant influence in shaping the policy environment for the collection and use of data.

An effective data governance structure must promote access that offers benefits to small businesses in particular, and the costs of adhering to the framework must not be prohibitive. Additionally, the realization of a single market for African data must be balanced with incentives for data localization, which is defined as a mandatory administrative or legal requirement indirectly or directly stipulating that data be stored and processed, non-exclusively or exclusively, within a specified jurisdiction. There are aspects of data localization that both advance and detract from effective data governance; although data localization can enhance data privacy and security, it can also inhibit trans-border data flows and lead to various negative consequences attendant to such a slowdown.

2.4.1 *Principles*

To fulfill its objectives, the data governance framework should adhere to certain central principles, including (a) promoting an agile framework to allow for innovation and experimentation; (b) ensuring accountability of all stakeholders within the data life cycle; (c) establishing standards for data accuracy and quality; (d) developing protocols for the standardization of data, thereby underpinning data quality and enabling interoperability; (e) preserving transparency in the utilization of data; (f) enabling equitable access of public data to all data users; (g) securing non-prohibitive costs of compliance to regulations relating to data; (h) promoting competition in the use and reuse of data; and (i) seeing to it that data sharing at an international level, outside Africa, occurs in full compliance with the rules of Africa's data governance framework.

The data governance framework should be governed by the principle of light-touch regulation, allowing innovation and experimentation, while still possessing the agility to respond quickly to information and implement lessons learned. The large number of use cases for data are unknown, and a restrictive regulatory stance discourages the realization of their full potential. A conducive environment for innovation can be established through regulatory sandboxes, as well as by leveraging the global experiences of other countries as they implement their own frameworks. A

conductive framework should also be promoted by regulating data applications appropriately as they are introduced, while still maintaining a principle of widespread availability of public data.

Accountability is a critical component of data governance that should be emphasized when developing both national and regional data governance frameworks. A comprehensive data governance framework covers dimensions of accountability both within and across organizations. Thus, at the macro level, an appropriate data governance framework should contemplate organizational dimensions to guide appropriate design within organizations, while also recognizing aspects of accountability from both a domestic and cross-border perspective.

Within an organization, the framework should promote a holistic view of data governance, as well as integration of data governance practices across departments, by directly and indirectly involved individuals. For example, the establishment of a data council with responsibility for data governance and with representation across departments and at all levels of seniority could formalize the creation of data policies and procedures for implementation and enable effective observation and monitoring. Organizational data governance frameworks currently in place tend to relegate data governance functions to an information technology department, often resulting in ineffective, fragmented, and partial implementation. Building integrated data governance at the organizational level will help build trust among stakeholders.

Data quality standards ensure the accuracy of data, build trust in its use, and allow for consistency in its dissemination. The sheer volume of data produced and analyzed on a daily basis, as well as its exponential growth, underlines the importance of maintaining data quality standards. Uncertainty about the quality of data discourages its use and, if relied upon, may result in erroneous decisions. In the worst case, data can be misused for malicious intent. Quality should, therefore, be ascertained to ensure data are timely, accurate, complete, and consistent—and quality standards should be compatible with other existing rules and regulations, including but not limited to those pertaining to privacy and competition.

Data standardization and the establishment of data protocols are also key enabling factors for supporting an effective cross-border data governance infrastructure and a single market for data. Data standardization contributes to ensuring data accuracy and has important implications for

productivity by improving the efficiency of data processes and also encouraging usage. Grannis et al. (2019) investigated the impact of data validation and standardization on accuracy, finding that, in the case of healthcare records, standardization increased the accuracy of healthcare data. What's more, standardization improves the interoperability and portability of data. Interoperability refers to the ability to integrate data sets from different sources, while portability means the ability to transfer or share data without affecting their quality and content. Interoperability standards should be established within countries, with comprehensive coverage over sectors, geographies, and interests, and these standards ought to be underlain with a set of appropriate technical frameworks (such as interfaces for application programming) that individuals can leverage in promoting interoperability.

Transparency should be exercised in all stages of the data governance process. Data-related decisions and data processes should be communicated across all data users to ensure a clear understanding of data-handling processes and to allow users to know how their information is obtained and deployed. This will, in turn, build trust within the data governance process, encouraging users to participate within the framework and providing users with the necessary information to exercise their rights with regard to the availability and use of their data. Moreover, access to data should be provided on an equitable basis across all categories, including both private and public data, and this access should be universal and independent of data producers' and users' economic status or market power. Access costs should be low enough to allow users widespread participation, and gaps in the enabling infrastructure within countries and across the continent should be closed to promote more equitable basis. Further, the costs of compliance for participation in the data economy cannot be prohibitive, an especially relevant concern in the African context, where the vast majority of the private sector consists of small businesses with limited capacity—businesses whose active participation will require a conducive framework that promotes competition in the use and reuse of data and drives innovative activity. Finally, to ensure that the rights of African citizens are adequately protected by third parties, the data governance framework should ensure that data sharing at an international level is done with those countries and regions who comply fully with the rules established within the African data governance framework.

2.4.2 *Infrastructure*

Without an appropriate enabling environment, including physical and human capital infrastructure underpinning its implementation, a data governance framework cannot reach its potential. This framework also calls for institutions to secure the right cultural, legal, policy, regulatory, organizational, institutional, and technical environment to ensure that all data users can effectively and efficiently extract value from data, as well as enforcement mechanisms of national and regional data governance frameworks, where regulatory authorities enjoy administrative, agency, and financial autonomy, to ensure the security and privacy of data.

The appropriate enabling environment should invest in infrastructure that lowers data access costs while increasing its quality, with particular attention to strengthening each country's infrastructure while also addressing access gaps both within and across countries. Investment should target a rapid increase in the digitization of data, promote the sharing of currently existing high-value public data sets, improve access to data at the household level, and raise the quality of public internet connections. Additionally, the value of the enabling infrastructure should be established by quantifying the impact of increased digitization, data access, and use both within countries and across the continent. The enabling environment depends on established protocols and application programming interfaces (APIs) to promote the standardization and transfer of data both nationally and regionally. Finally, these frameworks, as well as their productive deployment, hinge on investment in a well-trained labor force.

2.5 CONCLUSION AND RECOMMENDATIONS

Data sharing can vastly improve living standards through improvements in productivity. Data are also non-rival and partially excludable, meaning they can be reused infinitely without degradation. With knowledge building upon knowledge, returns to the utilization of data could increase in scale; the efficient utilization of data is thus critical to productivity growth.

Free data exchange has spawned unprecedented opportunities to people across the globe, creating jobs and industries, facilitating increased mobility, and ultimately raising standards of living. Policymakers aspire to responsible and safe data use to improve the lives of the people they serve, while minimizing the misuse or exploitation of data. To this end, a clearly defined data governance framework integrated with a data strategy is

necessary to establish data sovereignty and buttress Africa's competitiveness and cross-country collaboration during the digital age. Implementing an effective data governance framework will preserve the availability, usability, integrity, and security of data across the continent—and such a framework will be both served and safeguarded by a developed data infrastructure, technical protocols, laws and regulations, and institutions suited to promoting the safe and trustworthy use of data while respecting privacy.

The central aim of a pan-African data governance infrastructure is to help bring about a single market for data, thereby allowing the creation, use, and reuse of data by individuals across the continent and spurring economic growth and development while still protecting the rights of data subjects. To fulfill its objectives, this framework should adhere to certain central underlying principles and operate within a light-touch and agile regulatory framework that encourages innovation. These principles include ensuring accountability, maintaining data accuracy and quality, and facilitating interoperability and standardization of data; but just as much, this framework must afford equitable access of data and keep the costs of compliance low, so as to promote competition.

Some key areas require additional research if the maximum utility of a data governance framework is to be enjoyed, such as (a) identifying and prioritizing key strategic interests across the continent that will benefit the most from the implementation of a data governance framework, as well as quantifying the value of the framework; (b) determining strategies to increase the pace of digitization of offline public data sources, while also increasing access to already digitized public data; and (c) mapping infrastructure gaps and cost-of-access disparities at the subnational and cross-country level and then resolving these inequities. Finally, further research must explore how to create a regulatory framework that best implements the principles of effective data governance. This task will include, for example, research to define an effective accountability framework both at the firm and national levels, as well as the development of protocols to allow data transfer and interoperability.

REFERENCES

- Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2019). *Too much data: Prices and inefficiencies in data markets* (NBER Working Paper Series). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w26296/w26296.pdf

- Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 126(1), 3–9. <https://softwarezen.me/wp-content/uploads/2018/01/datawisdom.pdf>
- Allhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, 25(sup1), 64–75. <https://doi.org/10.1080/12460125.2016.1187397>
- Beraja, M., Yang, D., & Yuchtman, N. (2020). Data intensive innovation and the state: evidence from AI firms in China. NBER Working Paper No. 27723.
- Desjardins, J. (2019, April 17). How much data is generated each day? *World Economic Forum*. <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-daycf4bddf29f/>
- Fourie, J. (2016). The long walk to economic freedom after apartheid, and the road ahead. *IDEAS*. <https://ideas.repec.org/p/sza/wpaper/wpapers267.html>
- Grannis, S., Xu, H., Vest, J., Kasthurirathne, S., Bo, N., Moscovitch, B., Torkzadeh, R., & Rising, J. (2019). The effect of data validation and standardization on patient matching accuracy. *Journal of the American Medical Informatics Association*, 26(5), 447–456. <https://doi.org/10.1093/jamia/ocy191>
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152. https://www.researchgate.net/publication/220426163_Designing_data_governance
- Kouper, I., Raymond, A., & Giroux, S. (2020). An exploratory study of research data governance in the U.S. *Open Information Science*, 2020(4), 122–142. <https://doi.org/10.1515/opis-2020-0010>
- Marr, B. (2015, February 25). A brief history of big data everyone should read. *World Economic Forum*. <https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyoneshould-read/>
- Micheli, M., Ponti, M., Craglia, M., & Suman, A. B. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2), 1–15. <https://doi.org/10.1177/2053951720948087>
- Otto, B. (2011). *A morphology of the organization of data governance*. European Conference on Information Systems (ECIS 2011 Proceedings). https://www.researchgate.net/publication/221407900_A_morphology_of_the_organisation_of_data_governance
- United Nations Congress on Trade and Development. (2021). *Trade and Development Report*. United Nations Congress on Trade and Development. <https://unctad.org/webflyer/tradeand-development-report-2021>
- Vidal, M. F., & Medine D. (2019). *Is data privacy good for business?* Consultive Group to Assist the Poor. https://www.cgap.org/sites/default/files/publications/2019_12_Focus_Note_Is_Data_Privacy_Good_for_Business.pdf

- Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all—A contingency approach to data governance. *ACM Journal of Data and Information Quality*, 1(1), 1–27. https://www.alexandria.unisg.ch/67793/1/a4-weber_external.pdf
- World Bank. (2021a). *World development report 2021: Data for better lives*. World Bank Group. <https://doi.org/10.1596/978-1-4648-1600-0>
- World Bank. (2021b). *Minimum data consumption: How much is needed to support online activities, and is it affordable?* World Bank Group. <http://hdl.handle.net/10986/35149>
- Zimmermann, K. A. (2017, September 6). History of computers: A brief timeline. *Live Science*. <https://www.livescience.com/20718-computer-history.html>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

