



HAL
open science

A Critical and Systemic Consideration of Data for Sustainable Development in Africa

Nyalleng Moorosi, Mamello Thinyane, Vukosi Marivate

► **To cite this version:**

Nyalleng Moorosi, Mamello Thinyane, Vukosi Marivate. A Critical and Systemic Consideration of Data for Sustainable Development in Africa. 14th International Conference on Social Implications of Computers in Developing Countries (ICT4D), May 2017, Yogyakarta, Indonesia. pp.232-241, 10.1007/978-3-319-59111-7_20 . hal-01650102

HAL Id: hal-01650102

<https://inria.hal.science/hal-01650102>

Submitted on 28 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A critical and systemic consideration of data for sustainable development in Africa

Nyalleng Moorosi¹, Mamello Thinyane², and Vukosi Marivate¹

¹ Council for Scientific and Industrial Research, Pretoria, South Africa
nmoorosi@csir.co.za, vmarivate@csir.co.za

² United Nations University Institute of Computing and Society, Macau
mamello@unu.edu

Abstract. The “data revolution for development” pundits tout data as representing an undeniable opportunity for transforming and improving societies through the deployment of data-centric development approaches. The critics on the other hand question the legitimacy of these claims made on the role to data to transform society and development work, in particular considering the numerous systemic and structural challenges faced by some of the least developed countries. In this paper we consider the real positioning and role of the data, and in particular Big Data, for sustainable development in Africa. We highlight three perspectives and dynamics associated with the data revolution for development and suggest that the real utilization of data for development in Africa can only be realized when other ecosystem factors are considered in tandem.

Keywords: Data Science, Internet Participation, Open Data, Big Data, Data for Development

1 Introduction

The resolution 70/1 of the United Nations general assembly, which articulates the 2030 Agenda for Sustainable Development, galvanized global action towards the achievement of the 17 development goals and the 169 specific targets (United Nations, 2016). Effective development action towards these goals is dependent on an accurate understanding of the social well-being and environmental phenomena under consideration, and this in turn is dependent on the effectiveness of the indicators framework, and the quality of the observed metrics data. While countries have relied on and utilized data, typically collated by the National Statistics Offices (NSOs) for social indicators monitoring to inform their development policies and action, the twenty-first century presents an opportunity for the transformation of the social indicators monitoring domain through the developments in Internet technologies and also through the recent advent of Big Data.

Notwithstanding the discussions (Letouz, 2012; WEF, 2012) of the potential role of Big Data for sustainable development, it is not only necessary to critically interrogate the underlying developmental mechanisms and pathways of

data for development, but also to consider the systemic positioning of Big Data within national data ecosystems, taking into account the country-specific factors and conditions. In this paper we consider these aspects of the data revolution for sustainable development from the context of countries in Africa. In section 2, we present the theoretical framing of data for development and also discuss the related domain of social indicators monitoring. The diffusion of innovation model as well as the critical theory of technology are adopted to highlight inherent dynamics in the utilization of data for development. Section 3 adopts an ecosystem perspective to discuss the factors, contextualized for Africa, that support and enhance effective utilization of data for development. Section 4 then considers the opportunities and potential for the use of data for sustainable development in Africa. A conclusion that wraps the discussion on the importance of the ecosystem perspective and critical engagement in data for development is presented in section 5.

2 Data for development

The formal conceptualization of Data for Development (D4D) shares theoretical framing with the broader concepts of Knowledge for Development (K4D) and the Information and Communication Technologies for Development (ICT4D). In these frameworks information and technologies are viewed as indispensable resources and tools that are at the disposal of individuals, communities and governments towards their development. The utilization of these resources towards development can further be enunciated through the more nuanced theories such as the Capabilities Approach, which recognizes the potential of resources as inputs towards individuals capabilities (Sen, 1999). The consideration of data for development is largely undertaken from two distinct yet related perspectives: the social indicators monitoring perspective, and the development perspective. The former, and perhaps the most prominent, recognizes the role of data to revolutionize the work of monitoring development phenomena and of collating development statistics (Letouz, 2012; SDSN, 2015). The latter sees opportunities for data to directly impact individuals and communities developmental imperatives. This latter perspective does not represent a new thinking or a revolution to the human development discourse but rather highlights an emphasis on a specific resource (i.e. data) and its consideration within developmental contexts. The former perspective holds potential to revolutionize social indicators monitoring through the introduction of new actors, new data sources and new tools.

The collation of social indicators needs to be understood as an enabler and a step towards better decision-making, policy and developmental action. The failure of better scientific evidence, insights, and knowledge, to translate into better decision-making and better policy making is bemoaned across the board, from researchers in public administration and policy, to stakeholders in social indicators monitoring. Cloete notes the lack of, usually assumed, definitive causal link between availability of better information and the resultant quality of the decisions and outcomes taken (Cloete, 2009). Similarly, Cobb and Rixford note that

having relevant data about a phenomenon does not directly induce the resultant appropriate action (Cobb and Rixford, 1998). In order for data to be effective, it must be part of larger plan of action wherein evidence-basing approach is widely adopted as the core of policy development and analysis. This gap between evidence and action is a challenge in both developing and developed countries, and it is not a factor of the availability of quality data (Segone, 2008; GAO, 1995). The reasons for this failure, which is also termed the "utilization problem", include: failure to create ownership among the stakeholders, ineffective strategies regarding communicating the evaluation findings and data, lack of understanding of the political context and ecosystem factors, and failure to link the findings and data to a definite follow up plan (Segone, 2008). The effectiveness of data, applied to social indicators data, has been shown to be improved when the indicators and the data are clearly associated with a policy outcome or a definite plan of action (E. Innes and Booher, 2000). It remains therefore that far from the challenges of effective policy and development action being about the lack of social indicators data, in actual fact systemic and structural factors play a larger role in affecting the effective translation of evidence and insights into policy and action.

2.1 Critical perspectives on data for development

The critical theory of technology recognizes technology solutions as being socially shaped and constructed and therefore of being able to be used for rationalizing power structures as well as for empowerment (Zheng and Stahl, 2011; Feenberg, 1991). Critical consideration of data for development therefore necessarily dismisses both the technology determinism and the instrumental rationality that typically accompanies the discussions on the potential for data to revolutionize development (Cecez-Kecmanovic, 2005). Three perspectives emanating from the theory of diffusion of innovation and critical theory of technology are hereafter highlighted to suggest further issues that should remain within the locus of considerations of data for development.

Diffusion of hype This perspective is informed from the hype phase within Roger's Diffusion of Innovations which is typically accompanied by over-inflation of the potential of technology and therefore the associated expectations (Everett, 1995). In the data for development literature and related work this is seen through the fetishization of data wherein data, and in particular Big data, is purported as the missing factor in development work. Best engages with this aspect by highlighting the engagement with statistics "as though they are magical, as though they are more than mere number as though they distill the complexity and confusion of reality into simple facts as facts we discover, not the numbers we create" (Best, 2012). The over-emphasis on the role of data and the presumed data revolution that should transform development work and the implementation of SDGs is not only a naive proposition, it is also a risky one that shifts the focus away for the ecosystem factors that need to be taken into consideration

for effective development work. Data and social indicators as tools that help understand the social well-being phenomena should remain ancillary to the core development agenda (Cobb and Rixford, 1998).

The tyranny of benevolent technocrats From the perspective of critical theory, social indicators evaluation exists in a political landscape where values, beliefs, norms and power are contested. Thus, social indicators monitoring carries the overtones wherein the ruling class, or corollary in the case of SDGs the developed world, imposes certain values on the rest of the society (Cobb and Rixford, 1998). This phenomenon and its numerous implications for the global power dynamics has been well articulated and enunciated by Thompson in his critical study of the role of Information and Communication Technologies (ICT) in not only advancing the interest of specific technocratic stakeholders, but also in normalizing a certain socio-political worldview (Thompson, 2003). Further, the top down emphasis on the role of data (and Big Data) for development by the international development funding agencies necessarily imposes an agenda on the developing countries (those receiving international funding) that is not informed and driven bottom-up by the country specific considerations. The outcome of this tyranny of data dynamic is that the obligation-side of social indicators monitoring becomes the more emphasized, at the expense of leveraging the interplay between the enjoyment-side and the obligation-side towards informing holistic development policy and action (Green, 2001).

Plateau of empowered productivity Roger identifies the final stage in the diffusion of innovations as the productive utilization of the technological innovations (Everett, 1995). This would represent the use of data in development activities, which is characterized by : a clear understanding of the role and positioning of data within development activities; a context-sensitive use of data within a holistic and systemic development framework; accurate and transparent data analytics and statistics; and accessible reporting and dissemination of data for the various development actors and stakeholders. It is critical that the use of data in development serves the primary role and agenda of development. When the characteristics identified above as well as other country-specific factors are taken into consideration, the potential for effective utilization of data for development is increased.

3 Data ecosystem considered

The use of data for development exists within complex multifaceted systems comprising multiple stakeholders, processes, frameworks, standards and protocols, as well as platforms and systems. Effective operationalization of data for development is dependent on mature and optimized data ecosystems. This section considers some of the ecosystem factors, considered from the context of Africa, that have an impact of the use of data for development.

3.1 Connectivity and data availability

Data and information have always been utilized to support economic and societal development, however the Internet revolution and the recent developments around Big Data and social media data have elevated the role and potential of data for transforming social indicators monitoring and development work (SDSN, 2015). We consider the availability of social media data and indirectly the availability of the supporting connectivity, in the context of Africa to explore this potential. This preliminary exploration is undertaken for twenty African countries, made up of four clusters of five countries each from: the high Human Development Index (HDI), medium HDI, low HDI, and the lowest HDI. One of the critical factors that affects the availability of relevant digital data for social indicators monitoring is the extent of connectivity and participation of individuals on the Internet. In Africa there are increasingly more people who are connected through mobile devices, however this connectivity does not directly imply connectivity to the Internet, which could be affected by affordability, bandwidth availability and individuals capability; nor does it imply active "prosumption" (i.e. production and consumption facilitated by Web 2.0 tools) of data online. The "Active Internet Users" metric gives an indication of the potential generation of digital data from the different countries. From the ITU world telecommunication/ICT indicators database of 2016, out of the twenty countries under consideration in this study, the highest active Internet users (as a percentage of the population) is 69% and the lowest is 2%, for Kenya and Niger respectively (ITU, 2016). As expected and shown in Figure 1, the more developed countries have higher numbers of active users compared to the least developed countries, at the mean of 46% ($s^2 = 10\%$) and 2% ($s^2 = 4\%$) respectively.

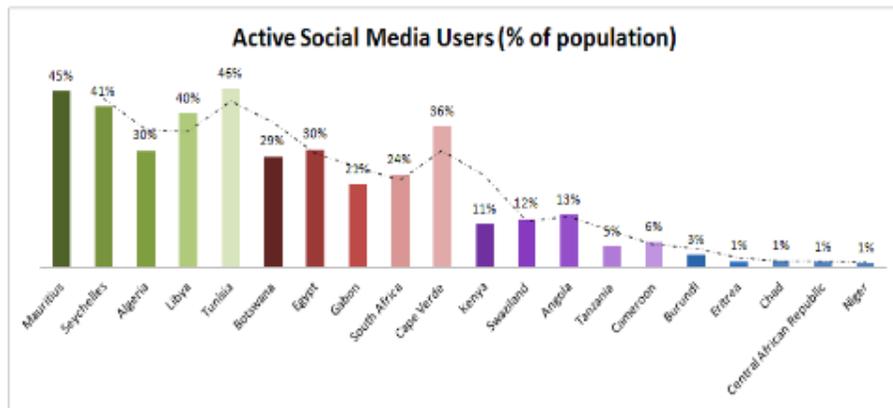


Fig. 1. Social Media Participation

The above metric highlights a phenomenon and a trend which is observable across various other metrics (e.g. International bandwidth per Internet user, percentage of adults accessing electronic services) and which has implications for development not only in Africa but across the world. The paradox of the assumed "Data revolution for development" is that the countries most in need (according to the widely accepted HDI model) of development are the same countries with minimal data repositories and relevant data sources, and in general those with data ecosystems that are not very mature.

Despite these challenges and limitations, it remains that the increasing real connectedness of individuals to the Internet, their greater participation in socio-economic activities, and the growing deployment of sensors and IoT devices, all represent a new opportunity in a form of new data sources that can contribute to informing the understanding of various social and environmental phenomena.

3.2 Privacy and governance

The second factor we consider is data security and privacy infrastructure. Unlike more advanced data markets such as the European Union and the Northern America, African countries are only starting to build policies and processes to regulate the use of data by entities within the data ecosystem (Borena et al., 2015). The 2014 African Union (AU) Convention Cyber Security and Personal Data Protection act is the first comprehensive attempt at developing an all-Africa cyber protection guideline (AU, 2014). However, this AU convention is yet to be rectified by member states. Thus practically, only a subset of African countries (i.e., Benin, Ghana, Tunisia, South Africa, Madagascar, and Gabon) have put into place legal frameworks that guide researchers and innovators in the Data Science space (Rick, 2015). Therefore for African states, the use of extensive datasets and especially those containing personal data may be difficult to justify.

3.3 Empowered and engaged citizenry

In the space of privacy and information ownership, most laws are designed to protect the use of personal data (e.g. name, demographics, and health data). However, in data mining and data science studies, publicly available information can be combined and processed in ways that reveal of the identity of the owner of that information. For instance users may choose to have anonymous identities on social media and may not be aware that the individuals in their network, their posts and their geographic location can reveal their identities. Thus, it is imperative that the individuals whose data is used in developing systems be educated about what can happen to their 'harmless' data such as their social media network. Citizens should be aware of the dangers of disclosing information as well as their rights in cases where their data is misused. Figure 2 provides a guide on the openness of information based on the visibility levels that users choose.

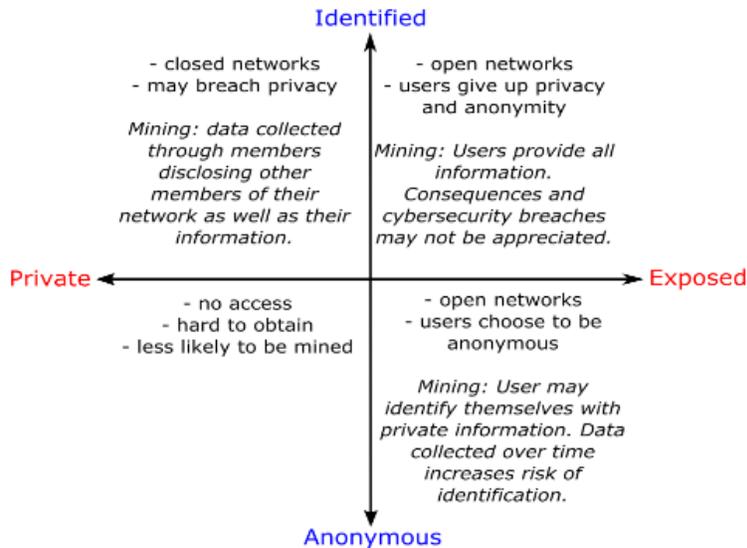


Fig. 2. Privacy in Social Media (Moorosi and Marivate, 2015)

3.4 Other Considerations

The following are other factors within the data ecosystem that are important for consideration.

1. **Big Data Skills:** There is currently a global shortage of individuals with Big Data skills. With competition from leading international technology companies such as Google and Facebook, all institutions are finding it difficult to recruit/retain enough staff to solve the problems pertinent to them.
2. **Computing infrastructure:** Technologically, deploying large scale Big Data infrastructure can be prohibitively expensive for small enterprises in developing countries. To meet the minimal demands, techniques in distributed computing and the use of open source software can be deployed to a very effective scale. That said, for institutions running at smaller data sizes, the simple use of old standard computers connected together as a cluster can provide enough power to run complex machine learning and Big Data storage platforms. Additionally deploying technologies such as Hadoop, and Spark can make those same old machines sufficiently efficient while allowing for scaling and future upgrades to meet increasing demand and complexity.
3. **Transparency and Data Availability:** Data is not only a resource, it is also a commodity. In Big Data analytics, beyond personal data, governments and companies may not have the inclination to share data with other stake-

holders within the data ecosystem. Initiatives such as the Open Data Forum work to improve this access.

4 Leveraging big data opportunities for Africa

Big Data, when effectively utilized, stands to supplement the current social indicators monitoring systems with actionable knowledge and insights that have largely been derived from electronic source streams. This can be data that is collected passively from users as they undertake everyday activities online (data exhaust), information that is generated directly by the users online (online information), data that is collected from sensors and IoT devices (physical sensors), and data that is collected from the public (crowd-sourced data) (Letouz, 2012). While the availability of the underlying data is an important and a necessary factor towards effective utilization of big data in sustainable development, it is also important that there is sufficient will and intent from the stakeholders, as well as the capacity and resources to process the data (Letouz, 2012). The opportunity for utilization of Big Data for sustainable development is manifested when each of these factors: availability, intent and capacity, are in place within country.

Other opportunities can be observed where the Big Data stands to provide solutions to some of the long standing challenges in Africa. Due to the unstructured nature of African cities: high growth rate and a high number of informal settlements (Arku, 2009), the problem of quantifying populations and understanding the built environment of the cities can be a challenge. It is commonly the most deprived members of the society that are the least known by the state and therefore the least heard and served. Without a proper understanding of demographics, African leaders cannot optimally plan and distribute resources such as health, education, and energy. Additionally they are at a disadvantage when they try to respond to matters of disease breakouts and natural disasters. Economically, not understanding the population means not fully understanding the state of several economic indicators such as employment, productivity and purchasing power parity and thus an impairment in designing optimal solutions for the citizenry. It is because of the above-mentioned that we see the use of citizens as contributors to geo-spatial annotation as one of key opportunities. The associated technologies and methods have already been used successfully in systems ranging from earthquake sensing to urban management (Laituri and Kodrich, 2008; Song and Sun, 2010) in developed and developing countries.

Furthermore, to achieve the goals of annotating and counting using citizen as sensor methods, African data scientists can piggy-back on already existing applications such as NextDrop³, a system that alerts residents when public water taps are open, health apps such as Mom Connect⁴, Find-a-Med a⁵ and Smart-Health-

³ <https://nextdrop.co/contact.html>

⁴ <http://www.health.gov.za/index.php/about-mom-connect>

⁵ <http://telecoms.com/interview/find-a-med-how-mobile-tech-is-changing-nigerian-healthcare/>

App⁶, which is deployed in several Eastern and Southern African countries. In most of these apps, there is a core GIS or mapping technology that is built in so as to provide relevant and reliable recommendations to its users. The use of this combination of demographic and spatial data can easily be adopted into machine learning tools that can decipher further information through more statistical and analytical methodologies.

5 Conclusion

The Data Revolution for Sustainable Development represent a potential opportunity for countries to transform their indicators data ecosystem towards supporting the realization of the SDG goals and targets. Beyond the allure of fetishization of data, the tyrannical influence of the technocratic stakeholders, and the naive misuse of data analytics tools and instruments, lies a domain of effective utilization of data not only to drive national policy on development, but also to support development action at the micro, meso and macro levels of society. This effective utilization of data is only realizable when the full ecosystem factors and dynamics, which are specific and unique to individuals countries, are earnestly considered. These factors include: the overall ecosystem readiness and capacity, frameworks for ethical processing of data, measures for data security and privacy preservation, as well as data governance models.

While recognizing the potential for the Data Revolution for sustainable development, we similarly note the the inherent paradox that the countries that would stand to benefit the most from this data revolution (through being the least developed) are the same countries that lack the data ecosystem maturity to effect and maximize this opportunity. In Africa the benefits of data to transform development monitoring will accrue to different countries at different levels, however overall, we have noted the positive impact and role that data can play in advancing the 2030 agenda for sustainable development on the continent.

⁶ <http://www.itnewsafrika.com/2014/03/smart-app-to-help-fight-malaria/>

Bibliography

- Arku, G.: Rapidly growing african cities need to adopt smart growth policies to solve urban development concerns. *Urban Forum* 20(3), 253–270 (2009), <http://dx.doi.org/10.1007/s12132-009-9047-z>
- AU: African Union convention on cyber security and personal data protection. African Union, Addis Ababa, Ethiopia (2014)
- Best, J.: *Damned lies and statistics: Untangling numbers from the media, politicians, and activists*. Univ of California Press (2012)
- Borena, B., Belanger, F., Egigu, D.: Information privacy protection practices in africa: A review through the lens of critical social theory. In: *System Sciences (HICSS), 2015 48th Hawaii International Conference on*. pp. 3490–3497. IEEE (2015)
- Cecez-Kecmanovic, D.: Basic assumptions of the critical research perspectives in information systems. *Handbook of critical information systems research: Theory and application* pp. 19–46 (2005)
- Cloete, F.: Evidence-based policy analysis in south africa: Critical assessment of the emerging government-wide monitoring and evaluation system. *Journal of Public Administration* 44(2), 293–311 (2009)
- Cobb, C.W., Rixford, C.: *Lessons learned from the history of social indicators, vol. 1. Redefining Progress* San Francisco (1998)
- E. Innes, J., Booher, D.E.: Indicators for sustainable communities: a strategy building on complexity theory and distributed intelligence. *Planning theory & practice* 1(2), 173–186 (2000)
- Everett, M.R.: *Diffusion of innovations*. New York (1995), <http://www.nehudlit.ru/books/detail8765.html>
- Feenberg, A.: *Critical theory of technology*. Oxford University Press New York (1991)
- GAO: *Programme evaluation: Improving the flow of information to the congress*. General Accounting Office (1995)
- Green, M.: What we talk about when we talk about indicators: current approaches to human rights measurement. *Human Rights Quarterly* 23(4), 1062–1097 (2001)
- ITU: *World telecommunication/ICT indicators database*. International Telecommunication Union. Geneva, Switzerland (2016)
- Laituri, M., Kodrich, K.: On line disaster response community: People as sensors of high magnitude disasters using internet gis. *Sensors* 8(5), 3037–3055 (2008)
- Letouz, E.: *Thoughts on Big Data and the SDGs*. United Nations Sustainable Development (2012)
- Moorosi, N., Marivate, V.: Privacy in mining crime data from social media: A south african perspective. In: *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec)*. pp. 171–175. IEEE (2015)
- Rick, S.: *Privacy security law report*. PVL R 1065 (2015)

- SDSN: Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development. Sustainable Development Solutions Network, New York, USA (2015)
- Segone, M.: Bridging the gap. the role of monitoring and evaluation in evidence-based policy making. (2008)
- Sen, A.: Development as freedom. Oxford University Press (1999)
- Song, W., Sun, G.: The role of mobile volunteered geographic information in urban management. In: Geoinformatics, 2010 18th International Conference on. pp. 1–5. IEEE (2010)
- Thompson, M.: Ict, power, and developmental discourse: A critical analysis. In: Global and organizational discourse about information technology, pp. 347–373. Springer (2003)
- United Nations: The Sustainable Development Goals Report 2016. Tech. rep., United Nations, New York (2016), <http://www.un.org/publications>
- WEF: Big Data, Big Impact: New Possibilities for International Development. World Economic Forum, Geneva, Switzerland (2012)
- Zheng, Y., Stahl, B.C.: Technology, capabilities and critical perspectives: what can critical theory contribute to sens capability approach? Ethics and Information Technology 13(2), 69–80 (2011)