Democratisation of data in an emerging economy with focus on Africa

By Obileye Olatunbosun Matthew (ORCID – 0000-0002-1200-0994)

Abstract

Data in an emerging economy are gradually being treated as an important resource, especially in commercial networks. Research data that are public funded are sometimes treated as personal data by many researchers in an emerging economy. In their study, David and William (2007) concentrated their efforts on democratisation of data by exploring the nature of information providers, the services provided, and the services provided to the affected communities but did not consider concerns associated with the process of data democratisation by the communities in an emerging economy. This paper looks at the democratisation of data, championed by Communication Unit of a research institute – International Institute of Tropical Agriculture (IITA), with presence in 35 countries within Africa (a continent filled with emerging economies), using a mini survey conducted, experience gathered in the process data democratisation as well as data generated in the course of data advocacy, data evangelism, and interviews. Massive data is being churned out in Africa, but many custodians of data are less interested in opening the data for public use. There is a new crop of researchers that see data democratisation as the future of wealth creation. This paper reviews the potentials of data democratisation in Africa and the challenges that come with it. Data democratisation in the paper is limited to scientific research data with little reference to business or commercial data.

Keywords: democratisation, FAIR data, advocacy, start-up, innovations

Introduction

Although research funding is gaining more attention in emerging economies like Brazil but bedevilled by poor financial management and institution (Minerva, 2005), the situation is yet to reach optimum level in Africa where only few of the countries are funding research. However, for not-for-profit organizations that are focused at eradicating hunger and poverty in Africa which also depends on international donors to fund their research, it is not expected that this will be the case.

Research data holds the key to new innovations. Such data, when made public could be the source for start-ups to scale up their businesses. In another clime, data made available to the public with appropriate license may be the catalyst that would improve lives in the impoverished countries with focus more on Africa.

Data democratisation makes data generated from public funds to be made available to the public in a digital format with no restriction for research re-use or business purpose (Bernard, 2017). Data democratisation ensures that more people with different data skills such as data engineering, data dive, data analysis, data curation, and other data-related fields such as artificial intelligence, machine learning and so on, can use available data to create opportunities for the society.

However, if the data itself is inconsistent from the source, the ripple effect of such untrusted data on the various secondary usage and output may be severe. In agriculture, for instance, crop breeding datasets based on specific soil sample under a controlled condition would be useful to the public provided the details of the soil biotic and abiotic factors are detailed in the data along with the seed type and prevailing atmospheric condition. The standard of measure, e.g., metric standard, must be well documented with a data dictionary from a defined ontology included in the data. If some of these factors are missing, it may be difficult to get the same expected yield if the data is used to prepare a commercial farm. In fact, it may lead to profit loss for such a commercial farm owner.

Data that is democratised must be complete and obey the FAIR principles of findability, accessibility, interoperability, and reusability (Wilkinson, 2016). It must be easy to find. Today, Google research dataset search engine may be a good testing platform for data that are open and easily findable. Such data should also be accessible. By accessibility, we mean that the data should not just be available for a period but perpetually, perhaps, in a repository that is credible in terms of long preservation. A core trust seal of approval may be enough to guarantee that such repository will be available perpetually. It is expected that the available data should be interoperable with other data from different sources. It should be machine readable and must be in bulk. Open format should be the standard of the stored data. A list of open formats can be found at http://opendefinition.org/ofd/ and http://opendatahandbook.org/guide/en/appendices/file-formats/. In addition, such data must be complete and re-useable. Complete does not include ethical concerned information like personally identifiable information (PII), privacy and legal concerned data.

Resistance to democratisation of data

Various authors have written about resistance and hindrances to open access. The case of data democratisation is not an exception. In times past, an oral interview conducted in the midst of research indicated that many were not keen at making research data available to the public after the completion of projects. The concerns of many of the researchers could be valid especially in Africa where plagiarism is yet to be fully controlled due to either the absence of proper policy on it or the non-enforcement of the policy. It could be discouraging for a research effort that might have spanned some years to be accessed and used by another party without attributing credit to

the author of the data. This singular scenario is one of the major concerns and challenges of data democratisation in Africa.

In 2017, some scientists, working in Africa, raised the question of how risky and fearful it would be if research data are made open to the public while a project is still active during a workshop on open access. They cited the example of a research data which was published prior to the release of a research publication but was cannibalised and used by another researcher based in another continent for the individua publication without giving credit to the author of the data. The data in question was generated in Africa but the first user of this data, who is not the actual author of the data, is in South America. The scientists' opinion was that before data could be democratised, the first author must have made at least, a publication from such data.

The same fear was a major factor in the survey conducted where 60% of the respondents considered the need to protect new discoveries from data collected as a major concern when it comes to democratisation of data. For a business entity that survives based on public data such as a retail business, the business may not be willing to release its data for public access in the best interest of the business. It is a common practice, for instance, in West Africa, to see product imitation with little form of enforcement to contain the adulteration of the original product in the market.
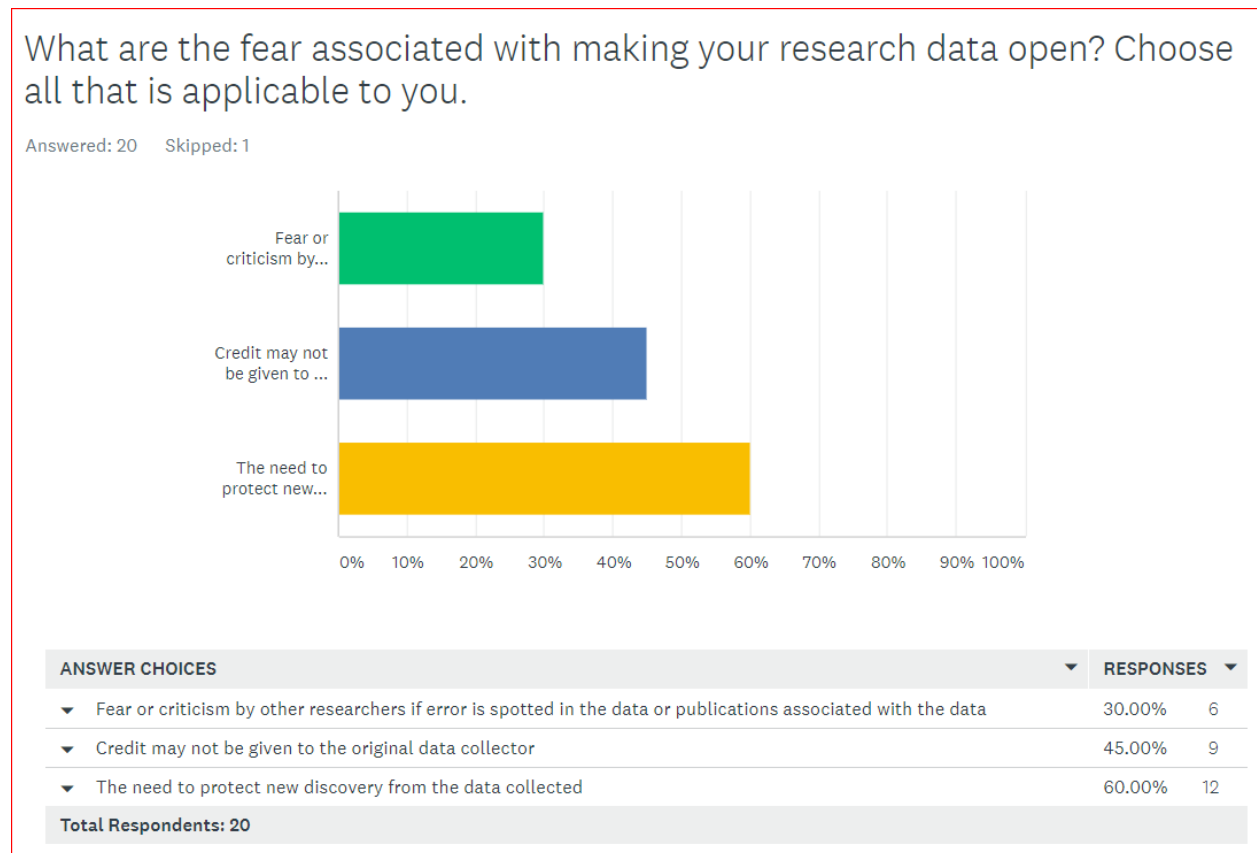
What are the fear associated with making your research data open? Choose all that is applicable to you.

Answered: 20   Skipped: 1

ANSWER CHOICES	RESPONSES	
Fear or criticism by other researchers if error is spotted in the data or publications associated with the data	30.00%	6
Credit may not be given to the original data collector	45.00%	9
The need to protect new discovery from the data collected	60.00%	12
Total Respondents: 20		

Figure 1: Ranking of fears of scientists with open access

The case of imitation of Coca Cola products in Nigeria by a former employee and branding it as Big Cola, Big Orange and Big MixFruit is a classic example of fears associated with data democratisation. The formula for preparing Coca Cola was accessed by some staff and the staff used the accessed data to create competition for Coca Cola by creating another brand with the same taste despite the patent right of Coca Cola. Although, Coca Cola took the case up in court, the indigenous firm won the case. Even though the data in this example is not open, being a trade secret, it could be part of the hidden concern associated with data democratisation in emerging economies, especially Africa.

New dawn in data world

1. Free data leading to new innovations

However, a small sample survey conducted among some young scientists from different countries in Africa in March 2019 indicated that 65% of the respondents were willing to research data they collected to be used for commercial purpose (see Figure 2). The population sample are well informed about the importance of open access and how it can transform the data ecosystem. Creation of innovative products and support for start-ups are two of the reasons for encouraging democratisation of data. Unlike the older generation, the new breed of data creators who responded to the questionnaire are seeing the bigger picture that comes with data that is free to the world. As much as this is good for the research or academic community, it may be a disaster to expose customer data to the public in a business environment. Banks and telecommunication industries are two examples of organizations that cannot afford to allow data democratisation to over-rule the customer's right to privacy, legal and ethical concerns.

Will you be happy if your data is used for commercial purpose without any monetary reward for you?
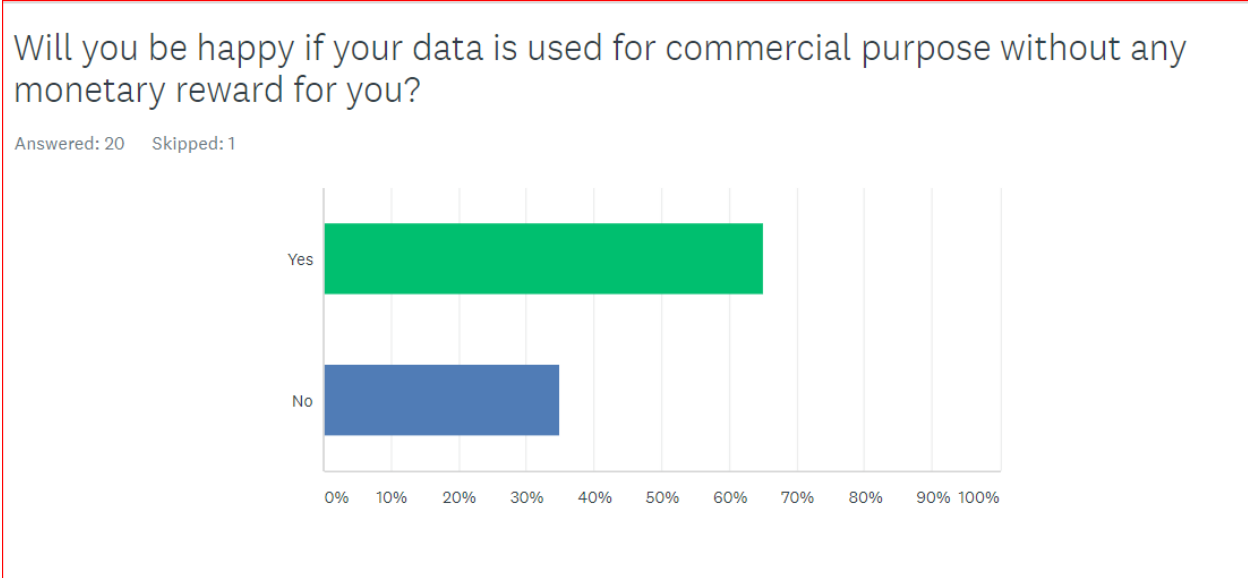
Answered: 20   Skipped: 1

Figure 2: Response to use of research data by 3rd party for commercial purpose

The implication of the new result from the small sample survey conducted among young scientists is that unrestricted access to data from research could help young scientists that do not have access to funds to make use of research data from others with access to fund and generate new conclusions, develop innovative technologies, and build on existing knowledge without reinventing the wheel. The concern about licence type may also be reduced to minimal if the data is open to the public for reuse. However, ethical, personal identifiable information, and privacy content within the raw data should be protected before the data is made public.

Acceptance of democratisation of data in today's business holds the door to innovations. Bernard (2017) believed making data available to teams within an organization will allow other highly skilled persons within the organization to develop out-of-the-box solutions that will enhance the life of the customers and add to the bottom line of the business. When businesses conduct customer satisfaction surveys, it is expected that the survey should be analysed and shared with internal stakeholders to improve the business process. Such a data may be transformed to tools by intelligent software resource within such organization if the data is democratized. It may even be used to develop a new product line. That implies that making data available within an organization can open more opportunities for such business.

2. Controlled freedom within democratised data

Another interesting fact from the survey showed that few of these respondents were willing to allow their research data to be accessed or used before publishing any article, journal, or paper on such data (See Figure 3). The result is interesting because of the demography of the respondents who are between 22 years and 35 years. This may be a good basis for supporting the freedom of access to data in Africa as a continent although this is a small sample data set because not all data content should be made public especially social science data that may contain confidential data of respondents. It is important to treat sensitive information from respondents such as address, full name, phone number, marital status, date of birth, number of family member(s), gender and geo-location that may be present in research data.
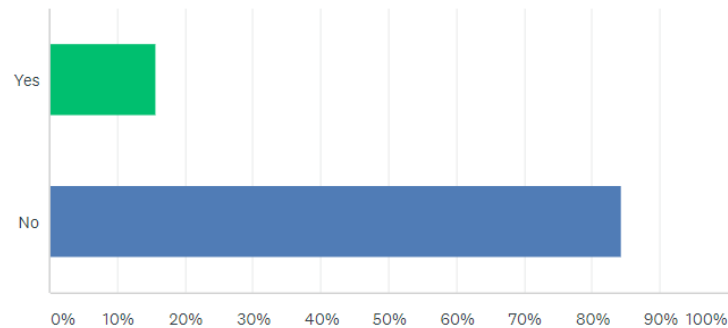
Figure 3: Willingness to make research data open before paper publication

Figure 1 also showed one of the concerns that would have made the respondent unwilling to make their data open as the data comes in – during the data collection phase of the project. The fear of the entire project being hijacked by remote users waiting for researchers to gather data and utilize it for a selfish purpose is real in Africa. It has been observed by research supervisors that students sometimes cook up data and analyse such data for research dissertation (Samuel & Omolara, 2009). In another case, there have been unconfirmed reports about data ownership, which has not been assigned a persistent identifier, hijacked by some selfish individuals who access data from open access repositories, and later assigned both ORCID and persistent identifier to such data in another repository within Africa. Such fearful unconfirmed cases may be reasons why the respondent may prefer to keep the research data on embargo until the end of data collection period. This may explain Figure 4 where majority of the respondents prefer to hide some important elements within the generated data before opening it up to the public.

Will you be willing to provide the entire raw data collected in your research or will prefer to hide some important elements before making the data open access?

Answered: 19   Skipped: 2
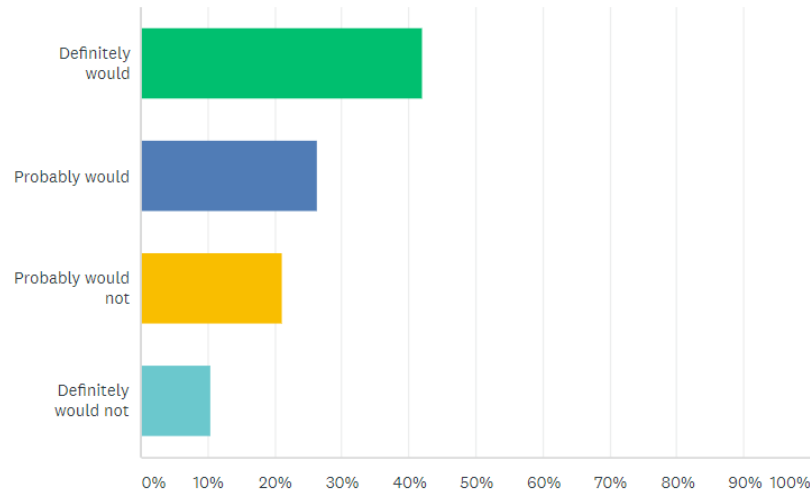
ANSWER CHOICES	RESPONSES	
Definitely would	42.11%	8
Probably would	26.32%	5
Probably would not	21.05%	4
Definitely would not	10.53%	2

Figure 4: Willingness to protect confidential data within data chart

Sustaining democratisation of data

Africa has a younger population compared to most developed economies (World Bank, 2019) which makes this paper and study relevant to the continent. To sustain democratization of data by this new generation of data generators, there is a need to provide the enabling ground and support such as funding, development of policies that will protect data democratisation in the continent while ensuring that privacy laws are considered for all data made public (See Figure 5). It is easier when it comes to a business entity to set aside funds for data generation but funding for research data within public or non-profit institutions and research centres in Africa is dwindling compared to the past, partly due to US government cuts in funding for most projects in its departments like USAID.
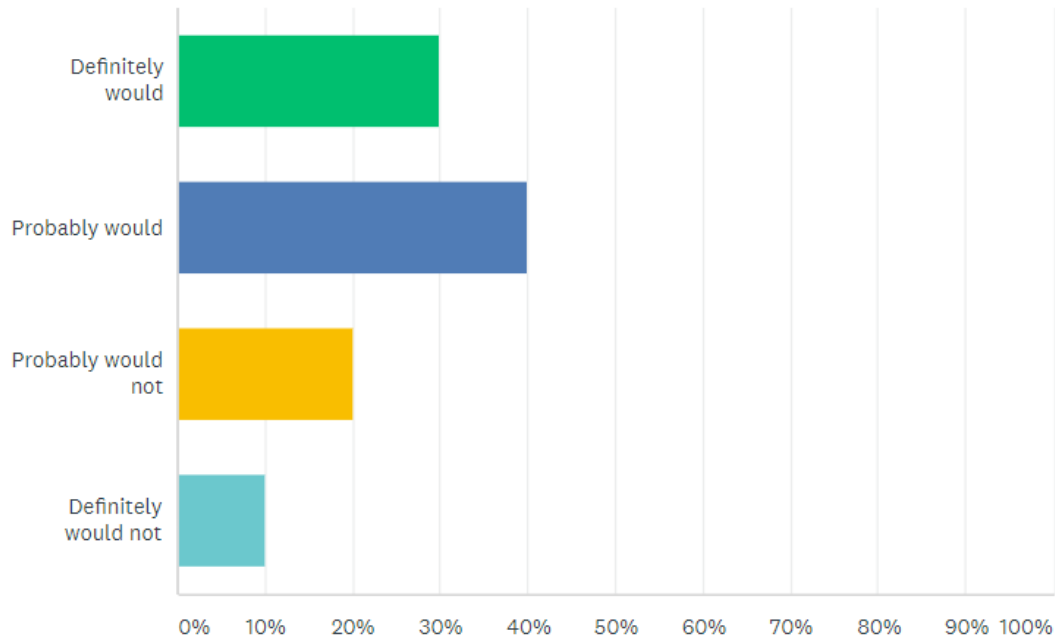
Figure 5: Chart showing the need for support to ensure data democratisation sustainability

Sustainability of democratisation of data is key to lift Africa from hunger and gradually move it to a self-sufficient economy. It is the resource that brings wealth to those that can transform data to opportunities. It creates more prospects for software developers and disruptive innovators while encouraging start-ups to develop technologies to open more opportunities. In agriculture, data is key to the development of agri-tech and the potentials of democratisation of data is truly unlimited.

References

Bernard Marr (2017) What is Data Democratization? A super Simple Explanation and The Key Pros and Cons, Forbes, https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/#52926d736013 [Online]

David S. Sawicki & William J. Craig (1996) The Democratization of Data: Bridging the Gap for Community Groups, Journal of the American Planning Association, 62:4, 512-523, DOI: 10.1080/01944369608975715

M.D. Wilkinson, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data3:160018 doi: 10.1038/sdata.2016.18.

World Bank (2019) "Population growth (annual%)" https://data.worldbank.org/indicator/SP.POP.GROW?end=2017&locations=ZG-ZF&start=1961&view=chart [Online]

Samuel C. Avemaria Utulu, Omolara Bolarinwa (2009) "Open access initiatives adoption by Nigerian academics", Library Review, Vol. 58 Issue: 9, pp. 660-669,

Sá, C. Minerva (2005) 43: 245. https://doi.org/10.1007/s11024-005-6474-4

Survey can be accessed from: https://www.surveymonkey.com/results/SM-XCBG5K7XV/ (Obileye Olatunbosun M., 2019)