Setting up data science research in Africa and engagement of stakeholders

AUTHORS:
Fati Murtala-Ibrahim[1] (iD)
Jibreel Jumare[2] (iD)
Manhattan Charurat[3] (iD)
Chenfeng Xiong[4] (iD)
Vivek Naranbhai[5] (iD)
Patrick Dakum[1] (iD)
Shirley Collie[6] (iD)
Waasila Jassat[7] (iD)
Gambo Aliyu[8] (iD)
Adetifa Ifedayo[9] (iD)
Alash'le Abimiku[2] (iD)

AFFILIATIONS:
[1]Institute of Human Virology Nigeria (IHVN), Abuja, Nigeria
[2]International Research Center of Excellence (IRCE), Institute of Human Virology, Abuja, Nigeria
[3]Institute of Human Virology, School of Medicine, University of Maryland, Baltimore, Maryland, USA
[4]Department of Civil and Environmental Engineering, College of Engineering, Villanova University, Philadelphia, Pennsylvania, USA
[5]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa
[6]Discovery Health, Johannesburg, South Africa
[7]National Institute for Communicable Diseases (NICD), Johannesburg, South Africa
[8]National Agency for the Control of AIDS (NACA), Abuja, Nigeria
[9]Nigeria Centre for Disease Control (NCDC), Abuja, Nigeria

CORRESPONDENCE TO:
Fati Murtala-Ibrahim

EMAIL:
fmurtalaibrahim@ihvnigeria.org

Significance:

Data science explores the use of big data to gain deeper insights and generate new knowledge and innovations which can lead to economic growth and sustainable development. However, setting up data science research comes with challenges. How we engage stakeholders is a major factor that determines success. This Commentary highlights important considerations for stakeholder engagement based on the experiences of investigators in a data science for health discovery project underway in Nigeria and South Africa. The perspectives presented will guide implementation in this relatively new but rapidly growing research domain.

Background

Health sciences research has been defined to include basic, clinical, and applied science on human health and well-being. It explores the determinants, prevention, detection, treatment, and management of diseases, and can be extended to data science research.[1] Setting up health sciences research in Africa will promote a strong health science industry as part of broader efforts to establish a robust research and development (R&D) environment, accelerating the emergence of knowledge-based economies that engender sustainable growth and development.[1] In a seminal 1990 report, the Commission on Health Research and Development stated that strengthening research capacity in low- and middle-income countries (LMICs) is "one of the most powerful, cost-effective and sustainable means of advancing health and development"[1]. Applying data science to health sciences research provides an opportunity to use large data sets generated in public health settings to gain deeper insight and generate new knowledge and innovations. It also provides better ways of implementing research to achieve greater health benefits, improving the economy of countries.[2-4] The Data Science for Health Discovery and Innovation in Africa (DS-I Africa) initiative aims to leverage data science technologies to transform biomedical and behavioural research. This initiative ultimately intends to develop solutions that would improve health for individuals and populations in Africa.[5] Likewise, the INFORM-Africa (Role of Data Streams in Informing Infection Dynamics in Africa) Research Hub partners with the governments, health facilities, industry, and communities in Nigeria and South Africa to advance data science in Africa by closing the gap in utilisation of big data and analytical capacity. The core objective is to strengthen the use of existing population-scale epidemiologic data sources as a cornerstone of future pandemic preparedness, using HIV and COVID-19 pandemics as examples.

Evolution of data science

Data science is an emerging and evolving discipline, especially in LMICs, and needs to be explored in sub-Saharan Africa to maximise the gains. Data science has been described by Beyene et al.[2] as an integrated interdisciplinary approach used to develop tools, templates, and processes to conduct complex analyses of big data sets. The authors attribute the slow evolution of data science in Africa and other resource-limited settings to a lack of well-trained data scientists.[2] 'Data science' as a term was created in the early 1960s and used to describe a discipline that supports the synthesis and interpretation of the large amount of data that had been generated over time, but it has evolved from statistics and data analysis to include computer science concepts like artificial intelligence, machine learning, and the Internet of Things.[2-4]

Stakeholder engagement

Data science research provides the opportunity for global collaboration amongst a wide range of expertise for maximum impact to be achieved. Tembo et al.[6] describe this collaboration in different ways depending on the region; for instance, in high-resource settings, it is known as 'patient and public involvement', 'engagement', or 'participation'. In LMICs, these approaches are termed 'community engagement', 'participation', and 'community engagement and involvement'.[6] In the same vein, the INFORM-Africa Research Hub has assembled experienced researchers with complementary expertise in big data analytics, quantum information processing, spatial statistics and analysis, genetics, computational biology, agent-based and data-driven modelling, clinical infectious diseases, infectious disease epidemiology, molecular virology, and geospatial analytics to address its research goal as outlined above.

Importance of stakeholders

The importance of stakeholders cannot be overemphasised. In most health-related research, stakeholder engagement and involvement can add value to the implementation of the research, in addition to building new knowledge and innovation.[7] It is important to make the stakeholders as broad and varied as reasonably possible, to engage them at the planning stage of the research when the priorities are being set, and to involve them in the design phase of the research project.[8] This helps to incorporate culturally acceptable norms into the research study proposed, ensures alignment with the priorities of the communities, supports the recruitment and retention of research participants, and facilitates overall implementation and dissemination of the research findings.[8-10] In summary, stakeholder involvement provides an opportunity for inclusion rather than exclusion.[11,12] Governments in Africa and most LMICs are important stakeholders for health-related data research as they own most health facilities at all levels of care. They also own most of the laboratories that generate results of clinical investigations. This makes the government an important stakeholder when big data generated by health facilities and laboratories is required for research. The INFORM-Africa Research Hub has benefitted from significant input

from all its stakeholders during grant proposal development, planning, and implementation. Its expanded and multidisciplinary stakeholders, including policymakers and communities of both countries, will play a significant role in disseminating its findings and products.

Challenges to data science research

In Africa, several challenges exist regarding access to big data and other aspects of data science research, including:

- lack of trained data scientists and inability to retain well-trained scientists (brain drain);
- limited infrastructure (facilities for curating research data, integrated Electronic Medical Record Systems, establishing national databases, electronic surveillance systems, national vital statistics repository, etc.);
- limited awareness of the value of data science research among researchers and research institutions;
- limited resources/funding for data science research;
- limited data sharing culture and opportunities;
- limited engagement of communities in research through community participatory research initiatives;
- limited training and availability of adequate provisions on regulatory/ethical guidelines for data science research;
- limited engagement of private health facilities and health insurance data sources; and
- poor clinical documentation, record keeping, and data management practices.

Standards in engaging stakeholders for maximum benefit

When standards that use the internationally recognised four foundational principles for scientific data management and stewardship – Findability, Accessibility, Interoperability, and Reusability (FAIR)[5] – are developed within the appropriate cultural context, they guide research teams to engage stakeholders and minimise the challenges often experienced. Some of the standards and guidance also provided by Tembo et al. include adopting the principles of power-sharing, building relationships, acknowledging diverse perspectives, reciprocity, and respecting different knowledge bases.[6,8] By partnering with government agencies, health data custodians, community gatekeepers, notable leaders in the scientific community, and research ethics boards in Nigeria and South Africa, INFORM-Africa has started these critical steps of engagement. We have incorporated representatives of our stakeholders in our standing committees, in addition to sharing and reviewing documents such as protocols, standard operating procedures and data sharing agreements, to ensure that elements of the FAIR principles are captured within the cultural context of these countries. Globally, engaging stakeholders and involving them in research efforts from conception to disseminating results will play a huge role in data science and changing policies.

Acknowledgements

We acknowledge the role of the DSI-Africa Consortium in making this research possible. The US National Institutes of Health provided the core funding for the DSI-Africa Consortium (more information is available at https://dsi-africa.org/). This research was specifically funded by the US National Institutes of Health (grant number 5U54TW012041-02). We also acknowledge the contributions of the INFORM-Africa project team, staff

and management of the Institute of Human Virology Nigeria (IHVN), University of Maryland, Baltimore (UMB), Villanova University, Centre for the AIDS Programme of Research in South Africa (CAPRISA), Nigeria Centre for Disease Control (NCDC), Discovery Health in South Africa, National Institute for Communicable Diseases (NICD) in South Africa, National Agency for the Control of AIDS (NACA), Nigeria National AIDS and STD Control Programme (NASCP), and all participants, patients, investigators, clinicians and personnel involved in generating the primary data used in the INFORM-Africa research project.

Author information

All the authors are part of the INFORM-Africa Research Study Group.

Competing interests

We have no competing interests to declare.

References

1. Wenham C, Wouters O, Jones C, Juma PA, Mijumbi-Deve RM, Sobngwi-Tambekou JL, et al. Measuring health science research and development in Africa: mapping the available data. Health Res Policy Sys. 2021;19(1), Art. #142. https://doi.org/10.1186/s12961-021-00778-y

2. Beyene J, Harrar SW, Altaye M, Astatkie T, Awoke T, Shkedy Z, et al. A roadmap for building data science capacity for health discovery and innovation in Africa. Front Public Health. 2021;9. https://doi.org/10.3389/fpubh.2021.710961

3. Cao L. Data science: A comprehensive overview. ACM Comput Surv. 2017;50(3), Art. #43. https://doi.org/10.1145/3076253

4. Navarro FCP, Mohsen H, Yan C, Li S, Gu M, Meyerson W, et al. Genomics and data science: An application within an umbrella. Genome Biol. 2019;20(1), Art. #109. https://doi.org/10.1186/s13059-019-1724-1

5. US National Institutes of Health (NIH). National Institutes of Health strategic plan for data science 2018 [document on the Internet]. c2018 [cited 2022 Sep 05]. Available from: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

6. Tembo D, Hickey G, Montenegro C, Chandler D, Nelson E, Porter K, et al. Effective engagement and involvement with community stakeholders in the co-production of global health research. BMJ. 2021;372, Art. #178. https://doi.org/10.1136/bmj.n178

7. Staunton C, Tindana P, Hendricks M, Moodley K. Rules of engagement: Perspectives on stakeholder engagement for genomic biobanking research in South Africa. BMC Med Ethics. 2018;19(1), Art. #13. https://doi.org/10.1186/s12910-018-0252-y

8. Goodman MS, Sanders Thompson VL. The science of stakeholder engagement in research: Classification, implementation, and evaluation. Transl Behav Med. 2017;7(3):486–491. https://doi.org/10.1007/s13142-017-0495-z

9. Musesengwa R, Chimbari MJ, Mukaratirwa S. Initiating community engagement in an ecohealth research project in southern Africa. Infect Dis Poverty. 2017;6(1), Art. #22. https://doi.org/10.1186/s40249-016-0231-9

10. Hinchcliff R, Greenfield D, Braithwaite J. Is it worth engaging in multi-stakeholder health services research collaborations? Reflections on key benefits, challenges and enabling mechanisms. Int J Qual Health Care. 2014;26(2):124–128. https://doi.org/10.1093/intqhc/mzu009

11. Boaz A, Hanney S, Borst R, O'Shea A, Kok M. How to engage stakeholders in research: Design principles to support improvement. Health Res Policy Sys. 2018;16(1), Art. #60. https://doi.org/10.1186/s12961-018-0337-6

12. Laird Y, Manner J, Baldwin L, Hunter R, McAteer J, Rodgers S, et al. Stakeholders' experiences of the public health research process: Time to change the system? Health Res Policy Sys. 2020;18(1), Art. #83. https://doi.org/10.1186/s12961-020-00599-5