Titre: Title:	Health data issues in Africa: time for digitization, standardization and harmonization
Auteurs: Authors:	Abdoelnaser Degoot, Ismaël Koné, Shakuntala Baichoo, Mercy Ngungu, Nzisa Liku, Judit Kumuthini, Joyce Nakatumba-Nabende, Foutse Khomh, & Bubacarr Bah
Date:	2025
Type:	Article de revue / Article
Référence: Citation:	Degoot, A., Koné, I., Baichoo, S., Ngungu, M., Liku, N., Kumuthini, J., Nakatumba-Nabende, J., Khomh, F., & Bah, B. (2025). Health data issues in Africa: time for digitization, standardization and harmonization [Commentaire ou lettre]. Nature Communications, 16(1), 5694 (4 pages). https://doi.org/10.1038/s41467-025-61104-6

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/66470/
Version:	Version officielle de l'éditeur / Published version Révisé par les pairs / Refereed
Conditions d'utilisation: Terms of Use:	Creative Commons Attribution-Utilisation non commerciale-Pas d'oeuvre dérivée 4.0 International / Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND)

Document publié chez l'éditeur officiel
Document issued by the official publisher

Titre de la revue: Journal Title:	Nature Communications (vol. 16, no. 1)
Maison d'édition: Publisher:	Nature Portfolio
URL officiel: Official URL:	https://doi.org/10.1038/s41467-025-61104-6
Mention légale: Legal notice:	This article is licensed under a Creative Commons Attribution-NonCommercialNoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article'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

Health data issues in Africa: time for digitization, standardization and harmonization

Abdoelnaser Degoot, Ismaël Koné, Shakuntala Baichoo, Mercy Ngungu, Nzisa Liku, Judit Kumuthini, Joyce Nakatumba-Nabende, Foutse Khomh & Bubacarr Bah

Check for updates

This commentary discusses health data challenges in Africa, focusing on digitization, standardization, and harmonization as key solutions. It highlights how addressing these foundational issues can enable AI and data science to transform healthcare systems across the continent.

Africa faces a dual health challenge with high infectious disease prevalence and a rapidly increasing burden of non-communicable diseases[1,2]. This significant disease burden is likely to worsen due to climate change, particularly through the emergence of climate-sensitive infectious diseases[3].

A beacon of hope emerges through Artificial Intelligence (AI) and data analytics technologies. Imagine AI-powered chatbots in remote villages offering basic health consultations in local languages or scanning vast medical databases to identify patterns for early and precise disease detection. These technologies could transform African healthcare through improved diagnostics, precision medicine, better public health systems, advanced drug discovery, patient monitoring, and remote consultation. The consequences of Africa not adopting these tools could perpetuate the already deplorable health systems, exacerbating negative socio-economic effects and widening existing health inequities[4].

Despite numerous health data sources and the increasing accessibility of low-cost data collection technologies, the continent still grapples with limited health data availability and poor data quality, creating a situation often referred to as "health data poverty"[4]. Although African health institutions have been collecting data for decades, most clinical data remains undigitized and siloed within institutions generating it[5]. This is due to a lack of comprehensive, standardized data collection processes. Many existing systems are program-specific and frequently updated to suit evolving needs. This situation presents a critical opportunity to establish proper data standards supporting the transition to widely adopted digital tools. AI and data analysis tools produce robust and reliable inferences when provided with high-quality data. In contrast, low-quality datasets compromise analytical validity—the "garbage in, garbage out" principle[6]. Critically, poor datasets can lead to misleading conclusions, potentially causing adverse outcomes. Furthermore, poor digitization severely impacts routine healthcare operations, planning, and evaluation in an increasingly digital world[7]. Figure 1 categorizes health data issues into four pillars. This commentary outlines how prioritizing digitization, standardization, and harmonization can address issues in the first two pillars: collection and management (Fig. 1). These technological solutions require implementation support from policy-makers and healthcare institutions to strengthen data infrastructure and unlock the transformative potential of data science in African healthcare systems. Addressing these two pillars will lay the foundation for effective data sharing and governance policies and large-scale AI implementation, both beyond this article's scope.

Data issues in the Africa healthcare system

Africa's health data issues range from limited availability and poor quality, to restricted accessibility, poor usability, lack of interoperability, and weak governance. Figure 1 summarizes these health data issues into four components reflecting the data science pipeline. We propose tackling health data ecosystem issues through digitization, standardization, and harmonization.

Data digitization. Significant analytics can only occur when data is digitized, making it easier to address data collection, management, and analytics issues (Fig. 1). Digitization facilitates fast access to comprehensive health data, fostering data-driven decision support within healthcare systems and enabling providers to track population health trends and respond rapidly to outbreaks using advanced analytics.

In Africa, healthcare digitization is primarily driven by electronic health records (EHRs). DHIS2 is the leading platform, used in 51 countries for health data management[8]. Electronic Community Health Information Systems (eCHIS) and portable data collection platforms supporting public health have also gained widespread adoption[9]. Other notable tools include REDCap for biomedical research and OpenMRS for patient records. Disease-specific programs like HIV/AIDS, malaria, and tuberculosis use custom EHRs, while private hospitals and research institutions often utilize their own systems. Despite these efforts, EHR adoption across Africa remains low[5], and the effectiveness of existing systems, measured by usage and data quality, is uncertain[10]. Inconsistent deployment due to lack of guidelines may affect healthcare facilities both within and across regions. To fully harness data science potential, Africa must pursue comprehensive digitization, supported by strategic investments in infrastructure (electricity, internet, computing equipment), maintenance costs, and capacity building. This includes developing standardized systems for data collection at the point of care. AI can assist by digitizing historical records and generating synthetic data for scenario modeling. Without digitizing health records, other processes like standardization and harmonization cannot be effectively implemented.
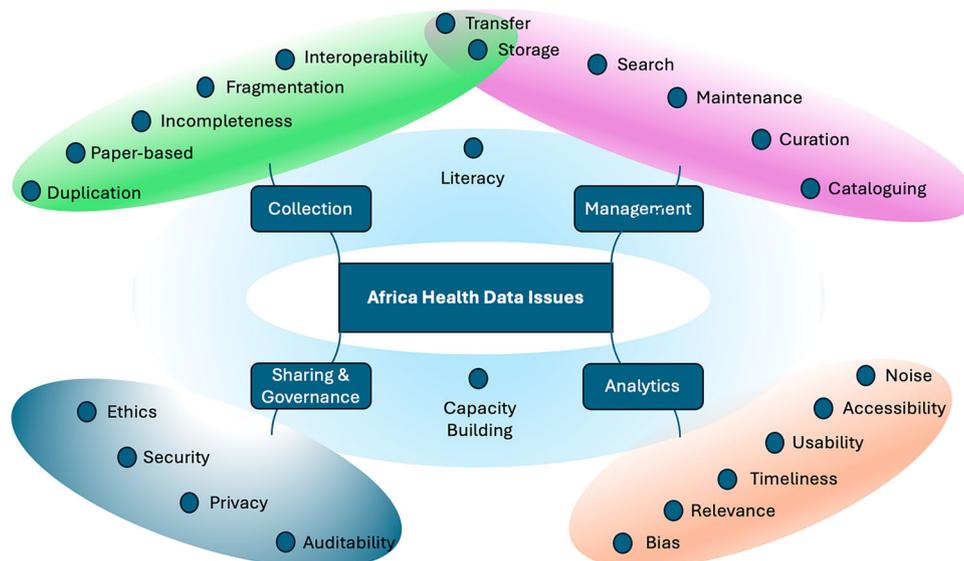
Comment

Fig. 1 | An illustration of key issues within the African health data ecosystem, varying across countries. Most challenges in the first two-pillars—collection and management—can be addressed through digitization, standardization, and harmonization. An added value of these technological solutions also lays the foundation for effectively tackling issues in the remaining pillars. Addressing these challenges collectively can greatly enhance the effectiveness and adoption of data science solutions in healthcare across the continent.

Data standardization. Data standardization ensures consistency in data format and structure[11], enabling uniform capture of variables across records and facilitating straightforward comparisons. For example, aggregating COVID-19 reports from multiple countries requires standardized case definitions and date formats to ensure accurate trend comparisons.

Achieving health data standardization requires both procedural efforts (shared guidelines and terminologies like SNOMED-CT and LOINC) and technological solutions (Common Data Elements and Models)[12,13]. Standardized data is particularly critical in African biomedical research due to the continent's linguistic and cultural diversity, ensuring reliable cross-country interpretation.

Standardization typically follows three models: (i) open source with community input (e.g., DHIS2, OpenMRS), allowing anyone to contribute; (ii) closed source with community feedback (e.g., RED-Cap), where developers rely on user feedback; and (iii) proprietary software with individual user feedback, common among private healthcare providers. Getting Africans actively engaged with these communities ensures the continent's perspectives shape emerging standards. Existing EHRs and research repositories often face standardization issues; specialized software for data cleaning, such as the cleanepi package in R[14], can expedite the post-collection curation process.

Data harmonization. Data harmonization integrates heterogeneous data from diverse datasets into cohesive datasets suitable for analytical studies. While standardization ensures uniform data formatting, harmonization focuses on making standardized yet heterogeneous data interoperable across platforms and regions, encompassing data linkage, fusion, and integration of multi-modal data. Harmonization consolidates large datasets, boosting statistical power in personalized medicine, disease research, and public health policy development. It ensures interoperability across diverse data sources—from genetic information to clinical records—and integrates data from various EHRs, supporting holistic health approaches like OneHealth. Without harmonization, health systems across Africa remain fragmented. For example, during the 2014–2016 Ebola outbreak, limited data sharing among affected countries hindered cross-border responses[15]. In South Africa, the HIV/AIDS data system isn't yet fully integrated with other disease monitoring programs, creating barriers for integrated patient care[16]. Efforts to harmonize African health data utilize open-source standards like FHIR[17], supported by initiatives such as RISLNET, OpenHIE, HELINA, and WHO Africa Observatory. While promising, these initiatives remain largely at pilot stages, with adoption lagging due to structural and infrastructural limitations.

What is the way forward?

Recommendations. We recommend that African ministries of health initiate or reinvigorate digital transformation of their health data systems by putting in place programs that:

- Invest in both physical and digital infrastructure and technologies, which are increasingly the backbone of functional health data systems.
- Establish and oversee the implementation of an agile digitization strategy, as digitization is demonstrably a core pillar in AI-powered health transformation.
- Adopt standards that ensure compatibility across systems and alignment with international standards. This will facilitate harmonization.
- Build human expertise and develop capacity in fields of study that will equip stakeholders with data science skills to build standardized, digitized, and interoperable health data systems. Realizing AI's potential in healthcare depends heavily on human expertise for implementation and sustainability.

Comment

To optimize the impact of the digital transformation across the African continent, ministries of health and regional health bodies (like the Africa CDC, WHO-Afro, WAHO, etc) should lead collaborative work on programs that:

- Foster interoperable systems enabling easy data exchange within and across regions, leveraging open standards and transparent APIs.
- Promote collaborative and inclusive solutions among public and private sector health stakeholders.
- Establish policy, legal, and regulatory frameworks aligned with national and regional laws, ensuring ethical data use, privacy, and security. Without effective data governance, even well-organized data systems risk becoming isolated silos, unsuitable for collaborative sharing.

In particular, regional health bodies should lead policy development that supports programs promoting interoperable systems with open standards and transparent APIs to enable seamless data exchange within and across regions.

Conclusion

The current state of health data digitization in Africa is limited, with standardization being inconsistent and harmonization nearly non-existent. This represents an opportunity to leapfrog directly to reliable AI adoption before fragmented approaches become entrenched. By implementing recommended actions, African nations can significantly increase standardized and interoperable digital health data. This shift could unlock data science's transformative potential, revolutionizing healthcare delivery, disease management, patient care, and public health policies continent-wide.

If Africa fails to embrace data science, the consequences will be severe. The continent risks exacerbating existing health challenges, stifling innovative research, and impairing decision-making processes, with profound economic implications—remaining trapped in persistent "health data poverty."

Abdoelnaser Degoot [1] ✉, Ismaël Koné [1,2] ✉,
Shakuntala Baichoo[3], Mercy Ngungu[4], Nzisa Liku[5], Judit Kumuthini[6],
Joyce Nakatumba-Nabende [7], Foutse Khomh [8] & Bubacarr Bah[1]
[1]Data Science, Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, Banjul, The Gambia. [2]Unité de Recherche et d'Expertise Numérique, Université Virtuelle de Côte d'Ivoire, Abidjan, Côte d'Ivoire. [3]Department of Digital Technologies, Faculty of Information, Communication and Digital Technologies, University of Mauritius, Réduit, Mauritius. [4]Impact Centre, Human Sciences Research Council, Pretoria, Gauteng, South Africa. [5]Public Health and Digital Health Expert, Nairobi, Kenya. [6]African Biobank and Longitudinal Epidemiological Ecosystem, Ibadan, Nigeria. [7]Department of Computer Science, School of Computing and Information Technology, Makerere University, Kampala, Uganda. [8]Mila–Quebec AI Institute, Polytechnique Montréal, Montréal, QC, Canada.
✉e-mail: abdoelnaser-mahmood.degoot@lshtm.ac.uk;
ismael.kone@lshtm.ac.uk

References

1. Vollset, S. E. et al. Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting analysis for the global burden of disease study 2021. *Lancet* **403**, 2204–2256 (2024).
2. World Health Organization, Regional Office for Africa. *The state of health in the WHO African Region: An analysis of the status of health, health services and health systems in the context of the Sustainable Development Goals* (WHO Regional Office for Africa, 2018). Licence: CC BY-NC-SA 3.0 IGO.
3. Uwishema, O. et al. Impacts of environmental and climatic changes on future infectious diseases. *Int. J. Surg.* **109**, 167–170 (2023).
4. Ibrahim, H., Liu, X., Zariffa, N., Morris, A. D. & Denniston, A. K. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit. Health* **3**, e260–e265 (2021).
5. Musa, S. M. et al. Paucity of health data in Africa: an obstacle to digital health implementation and evidence-based practice. *Public Health Rev.* **44**, 1605821 (2023).
6. Mohammed, S. et al. The effects of data quality on machine learning performance on tabular data. *Inf. Syst.* **132**, 102549 (2025).
7. Jeilani, A. & Hussein, A. Impact of digital health technologies adoption on healthcare workers' performance and workload: perspective with doi and toe models. *BMC Health Serv. Res.* **25**(271) (2025).
8. DHIS2. District health information system 2. https://dhis2.org. Accessed: 2024-09-08.
9. Bogale, T. N. et al. Acceptability and use of the electronic community health information system and its determinants among health extension workers in Ethiopia: a retrospective cross-sectional observational study. *BMC Med. Inform. Decis. Mak.* **23**, 290 (2023).
10. Océane, J.M. et al. Digital health tools could boost efficiency in Africa health systems. https://www.mckinsey.com/industries/healthcare/our-insights/how-digital-tools-could-boost-efficiency-in-african-health-systems. Accessed: 2024-08-16.
11. Umberfield, E., Bowie, J., Kanter, A., Dixon, B. & Tallman, E. Chapter 10—standardizing health care data across an enterprise (2023).
12. Sciences, T. O. H. D. & Informatics. Standardized data: The OMOP common data model. https://www.ohdsi.org/data-standardization. Accessed: 2024-08-19.
13. Rolland, B. et al. Toward rigorous data harmonization in cancer epidemiology research: one approach. *Am. J. Epidemiol.* **182**, 1033–1038 (2015).
14. Mané, K., Degoot, A., Ahadzie, B., Mohammed, N. & Bah, B. *cleanepi: Clean and Standardize Epidemiological Data* https://doi.org/10.5281/zenodo.11473985, https://epiverse-trace.github.io/cleanepi/ (2024).
15. WHO. Ebola response report 2016. https://web.archive.org/web/20190722235743/https://who.insomnation.com/sites/default/files/Images/ebola-response-report-2016.pdf. Accessed: 2024-09-10.
16. Clouse, K., Phillips, T. & Myer, L. Understanding data sources to measure patient retention in hiv care in sub-Saharan Africa. *Int. health* **9**, 203–205 (2017).
17. International, H. L. S. Fhir release 4 (v4.0.1): Overview. https://www.hl7.org/fhir/overview.html. Accessed: 2024-09-16.

Author contributions

The NIH team and Co-Chairs came up with the topic of the article as part of a Nature collection of health data science in Africa. They selected B.B. to lead a group of authors. B.B. selected a co-lead, F.K., and some more authors to increase the diversity in the group of authors in terms of expertise, gender, and geography. B.B., F.K., S.B., M.N., N.L. and J.N. all participated in the original discussions about the structure and content of the paper, and subsequently, including J.K., gave feedback to the text put together mainly by A.D. and I.K.

Competing interests

The authors declare no competing interests.

Ethics approval

As a commentary article and without the involvement of animals or humans, an ethics approval was not needed. All authors are from the Global South.

Comment