Research Article

Responsible Governance of Genomics Data and Biospecimens in the Context of Broad Consent: Experiences of a Pioneering Access Committee in Africa

Ahmed Rebai[1], Akin Abayomi[2], Pamela Andanda[3], Daima Bukini[4], Robyn Kerr[5,6], Kobus Herbst[7], Jenniffer Mabuka[8,9], Rosalyn Wamuyu[10], Collet Dandara[11]

1. Laboratory of Molecular and Cellular Screening Processes, Computational Biology group, Centre of Biotechnology of Sfax (CBS), Sfax, Tunisia; 2. Division of Hematopathology, Faculty of Medicine and Health Sciences, Tygerberg Hospital, University of Stellenbosch, South Africa; 3. School of Law, University of the Witwatersrand, Johannesburg, South Africa; 4. Sickle Cell Programme, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 5. Division of Human Genetics, National Health Laboratory Service, Johannesburg, South Africa; 6. School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 7. Africa Health Research Institute, Durban, South Africa; 8. University of Washington, United States; 9. Independent Consultant, Science for Africa Foundation, Nairobi, Kenya; 10. Human Heredity & Health in Africa (H3Africa) Programme, Science for Africa Foundation, Nairobi, Kenya; 11. SAMRC/UCT Platform for Pharmacogenomics Research and Translation Unit, South African Medical Research Council, Cape Town, South Africa

International collaboration in genomic research is gaining momentum in African countries and is often supported by external funding. Over the last decade there has been an increased interest in African genomic data. The contribution of this rich data resource in understanding diseases predominant in both African and global populations has been limited to date. Although There has been some non-governmental funding dedicated to the advancement of genomic research and innovation by African-based and African-led research groups, but the impact of these initiatives is hard to quantify. However, there is now opportunity for the global research community to leverage decades of genomic data and biospecimens originating from African populations. The experience we describe in this paper is of an access governance framework established under the Human, Heredity, and Health in Africa (H3A) consortium, given the task of managing wider access to the data and biospecimen resources collected via its various projects. The function of the Data and Biospecimen Access Committee (DBAC) is to facilitate the advancement of medicine and health, whilst fostering the development bioinformatics capabilities at Africa-based institutions or regional hubs. Our collective experiences and lessons learned as a committee provide examples of nuanced considerations when evaluating access to African data. The committee was semi-autonomous in its establishment and has independence in decision-making. The DBAC continually advocates for responsible use of genomic data and biospecimens that were obtained from African research

participants, under broad consent, by primary researchers who no longer have oversight over future use of these resources.

Corresponding author: Ahmed Rebai, ahmed.rebai@cbs.rnrt.tn

Abbreviations

- **DAR:** Data Access Request
- **DAC:** Data Access Committee

Introduction

With Africa entering the genetic characterization space, and the resultant generation of whole exome or genome sequence data for thousands of individuals, facilitating responsible access to genomic data is an ethical and scientific imperative. Generating genomic data from African populations has a significant benefit to the broader scientific community due to extensive genetic diversity observed in African genomes. Whilst it is in the best interests of Africans and their health that genomic data be accessible to researchers and health care professionals across the global community, access needs to be both facilitated and controlled.

Access to genomics data can be regulated via three possible approaches:

- **Unrestricted or public access:** data is completely open and accessible for consultation and public download. This kind of access does not allow traceability and monitoring of the intended use of the data.
- **Registered access:** requires the registration of the user to the repository before being able to download the data. This guarantees traceability and monitoring of the use of the data.
- **Controlled access:** is the most stringent data access approach and requires credentialed users to apply for access to data through an intermediary body which evaluates requests and allows access if a request complies with stated prerequisites.

For data repository, there is not a "one-size-fits-all" approach, it is possible to have any mixture of the three types of access, where some datasets/variables are open, some require simple registration, while others may need a data access request. In determining which of these approaches are suitable for the African context considerations such as ease of access to data/samples, subsequent benefit sharing, and vulnerability of relevant population groups, should be treated as major issues associated with data and sample sharing. There is a need for governance to ensure appropriate access and use of data, to avoid introducing data errors into systems and to obviate potential misuse of personal and sensitive information [1].

In controlled access, data requests (subsequently abbreviated as DARs) are evaluated by a Data Access Committee (DAC). A DAC is a body composed of individuals with diverse expertise, removed from the processes of obtaining biological samples and genomic data generation. The CIOMS guidelines [2] state that DACs should have "representation from the original setting" with diverse expertise that cover data management, ethics, analytics relevant research areas and a patient or data sharing advocate.

DACs are responsible for data release to external requestors based on guidelines and policies which have been defined by the parties that collected the samples, produced the data and funded the process, whilst respecting the regulations of the country or continental region from which the samples originated. DACs should be established within institutional and legal frameworks with clear lines of accountability, terms of reference and membership. DACs assess the ethical soundness and scientific feasibility of the DARs and evaluate the qualification of applicants to ensure they are *bona fide* researchers [3]. DACs can be either independent of the institution that produced the data, or institutional, each scenario presenting advantages and disadvantages [4]. It has been recognized that DACs play important roles in both promotion of data sharing and protection of all stakeholders (data subjects, communities, data producers, institutions where researchers are based, the country or even continent of affiliation) by encouraging secondary uses of data.

In this paper, we share the experience of the Human Heredity and Health for Africa (H3Africa) DBAC in assessing and evaluating DARs for the genomics data that have been generated under the H3Africa project. We discuss the ways in which the work of this (or similar) DBACs can be improved to achieve the goals of simultaneous stimulation of genomics data production in Africa, and the subsequent responsible wide sharing of these data.

The H3Africa DBAC

The H3Africa project

African genomics data have been very scarce and represent <2% of all genomics data available worldwide [5] [6]. To fill this gap, the H3Africa program was launched in 2012 with funds from the NIH (USA) and the Wellcome Trust (UK) in partnership with the African Society of Human Genetics. The H3Africa consortium facilitated fundamental research on genetic determinants of diseases on the African continent by developing human capacity, infrastructure (mainly for sequencing, biobanking and bioinformatics), resources and ethical guidelines, by funding 51 projects led by African scientists within the continent. Projects included population-based genomic studies of common non-communicable disorders, as well as infectious diseases. The program ended in 2022 and resulted in the release of 23 datasets (Table 1) and the collection of 23421 biospecimens from 16 studies [7].

Dataset identifier[a]	Data type[b]	Sample size (# participants)	Pop. group	Project name	# times requested	Requesters Country
EGAD..6295	SNP	10	Mali	NEEDI	3	UK, India, Nigeria
EGAD..1258	GA	973	South Africa	AWI-Gen	5	USA, UK India, Nigeria
EGAD..1996	GA	265	Burkina Faso, Kenya, Ghana, South Africa	AWI-Gen	6	USA, UK, Nigeria
EGAD..4557	GA	50	Benin	Malsic	17	USA, UK, SA, Nigeria
EGAD..6224	WES	314	Botswana Uganda	CAfGEN	7	Germany, UK, India, SA
EGAD..6418	WGS	100	South Africa	AWI-Gen	7	USA, UK, SA, Nigeria
EGAD..8577	WGS	410	South Africa	H3A chip TrypanoGEN	6	USA, SA, Nigeria
EGAD..4448	WGS	60	Burkina Faso Ghana	AWI-Gen	17	USA, SA Europe
EGAD..4220	WGS	41	Zambia	TrypanoGEN	17	USA, SA Europe
EGAD..4393	WGS	26	Cameroon	TrypanoGEN	17	USA, UK, SA, India, Sweden, Brazil
EGAD..4533	WGS	48	Botswana	CafGen	17	USA, UK, SA, India, Sweden, Brazil
EGAD..4316	WGS	24	Cameroon	H3AChip-Elsi	17	USA, UK, SA, Nigeria
EGAD..4334	WGS	50	Mali	NEEDI	17	USA, UK, SA, India, Sweden, Brazil, Nigeria

Dataset identifier[a]	Data type[b]	Sample size (# participants)	Pop. group	Project name	# times requested	Requesters Country
EGAD..4505	WGS	49	Nigeria	ACCME	16	USA, UK, SA, India, Sweden, Brazil
EGAD..5310	WGS	348	Zambia, Cameroon, Mali, Nigeria, Botswana, Benin, Burkina Faso Ghana.	H3A-ChipDesign	5	USA, Nigeria
EGAD..5076	WGS	233	Guinea, Côte D'Ivoire, DRC, Uganda.	TrypanoGEN	13	USA, UK, Germany,SA, India, Sweden, Brazil
EGAD..6425	Phenotype	12032	Ghana, Burkina Faso, Kenya, South Africa.	AWI-Gen	4	USA, UK, Nigeria
EGAD..6244	Phenotype/MG	196	South Africa	ReMAC	0	
EGAD..6581	Phenotype	171	South Africa	AWI-Gen	2	India, Nigeria
PRJEB40733	16S RNA	15		AWI-Gen	1	Nigeria
PRJEB37312	Shotgun microbiome	23		ReMAC	0	

Table 1. Summary of the datasets available in the H3Africa catalogue, accessible via the European Genome Archive (EGA) following request to the H3A-DBAC (EGA identifier EGAC00001000648). Included is information on how many times the respective dataset has been requested, plus the countries from which the request originated.

[a.] *Datasets are identified in the H3Africa catalogue by an identifier in the EGA database; all genome data (whole genome, whole exome, genotyping arrays) of H3Africa are identified by a code that starts with EGAD0000100xxxx, where only the last four numbers are unique.*

[b.] *GA: Genotyping array, WGS: Whole genome sequence, SNP: Single nucleotide polymorphism identified by Sanger sequencing of candidate genes*

The H3Africa DBAC

All genomic and phenotypic research data generated by the H3Africa projects were securely stored at the H3Africa Data Archive, under the H3ABioNet, a Pan African bioinformatics network for the consortium, and then submitted to the European Genome-Phenome Archive (EGA), an upfront requirement stipulated by the funders. The decision to use EGA was based on the lack of suitable infrastructure on the continent for data storage, and the long history and expertise of the EGA with respect to handling this type and amount of data. Data access is controlled by the H3Africa Data and Biospecimen Access Committee (referred to as H3A-DBAC), which is in accordance with the EGA guidelines that allow any organization, project, or funder that deposit data to create their own DAC for controlled access (Figure 1). Currently, 2432 active DACs operate under the EGA [8].

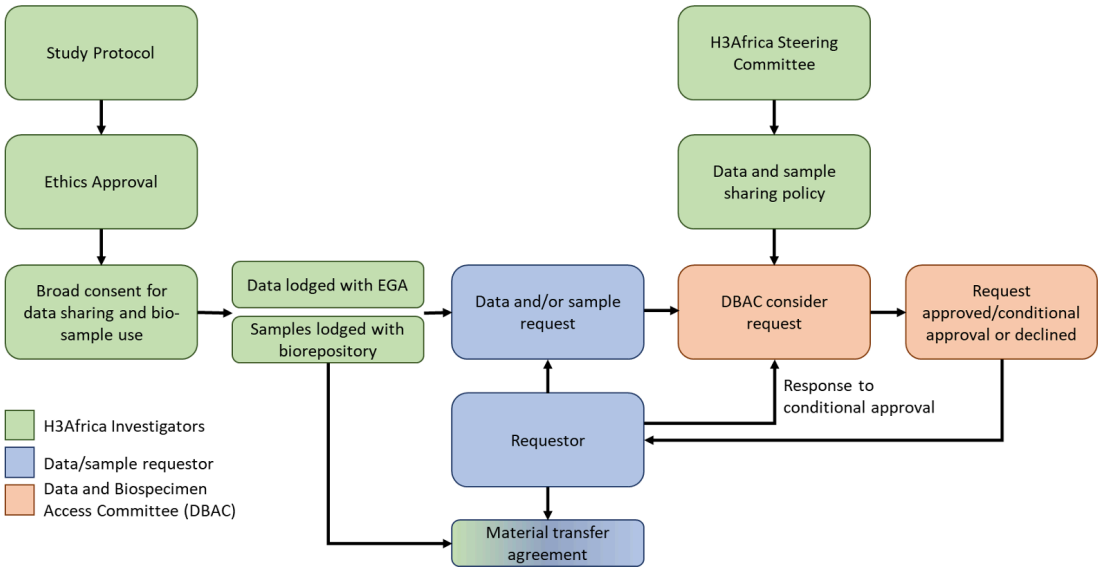

Figure 1. Data and Biospecimen Access committee request processing flowchart

The H3A-DBAC was formed in 2016, at the request of the H3A steering committee. The guidelines under which the DBAC should operate, and considerations for data access, were developed by a subcommittee of the consortium, the DBAC Working Group. There was subsequent input from the DBAC members toward refining certain guidelines for evaluating DARs. These guidelines are available in [9].

The H3A-DBAC is currently composed of 9 members from across the continent, with expertise in different fields: biobanking, data, genetics and genomics, ethics, law, patient advocacy, demography and epidemiology. The mandate of the committee is to "evaluate whether the requests conform to H3Africa policies and procedures including consistency of the proposed research use with the informed consent under which the data or biospecimen were collected and any other limitations stipulated by the submitting H3Africa investigators for each study." While guidelines were initially stipulated by the steering committee, the DBAC acts independently as it adjudicates on requests. The committee also provides continued feedback to the DBAC-Working Group on aspects that need to be reconsidered, as they learn from their activities.

Challenges faced during the set-up of the H3A-DBAC

The H3A-DBAC is, to the best of our knowledge, the first DAC that was set up to regulate controlled access to African genomics data produced on the continent. The context of data and biospecimen sharing in Africa includes difficulties associated with identifying appropriate participants, consenting them properly, and the preparation and storage of samples. In addition, processing samples, particularly genome characterization, is costly and the subsequent analysis comes with its own challenges. Thus, any sample and data sharing governance would need to address these realities.

The key issues that required significant discussion were

1. the requirement for the data requestor to return benefits to African populations from the intended secondary use of the data.
2. the requirement for the data requestor to demonstrate evidence of a collaboration with African scientist(s), which was stipulated as being mandatory for biospecimen requests and recommended for data requests [9]. The DBAC rationalized that this would promote capacity building on the continent through collaborative partnership.
3. access to data by commercial/non-profit organizations.

H3A DBAC: The Experience of the First Five Years

The first DAR was received by the DBAC in December 2018. As of end June 2023, a total of 28 DARs have been received and evaluated by the DBAC (Figure 2a).

Origin of the requests

Seven DARs (25%) were from private companies, 6 from Africa, while 13 were from USA-based organizations (Figure 2b).

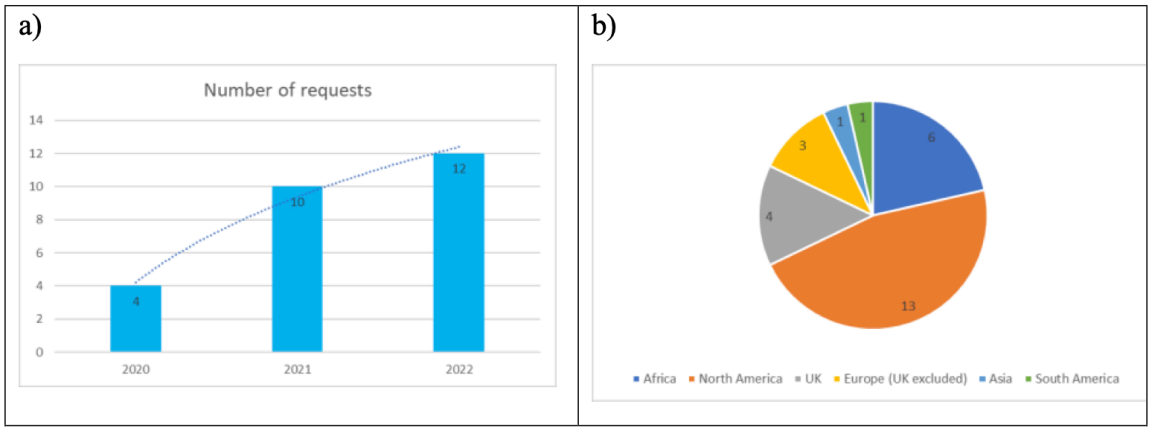

Figure 2. Number of requests received per year (a) and by geographic origin (b)

Time taken to first feedback on the DARs

When requests were made, their first contact is the H3A-DBAC secretariat, which quality checked the completeness of supporting documentation, after which the requests were forwarded to DBAC members for evaluation and decision. At the outset, the DBAC met quarterly, but this was changed since end 2022 to monthly meetings to facilitate quicker turnaround times. The average time taken by the DBAC to examine the DARs and issue a first reply was 69 days (20-210) and a median of 57 days.

Number of datasets requested

Most of the DARs (19/28, 67.8%) requested three datasets or more, while seven requested a single data set and two requested two datasets. Among the DARs requesting access to a single dataset, four concerned whole genome sequences (WGS) (see Table 1).

Decisions on the requests

The first decision of the DBAC for any request would be:

- ***Approval*** - 3 requests were approved upon first review.
- ***Conditional approval*** - this normally resulted from requests missing information that is required, or DBAC members seeking clarification of some issues. The requester would be asked to either make revision or supply the missing information and documents. The DBAC then carried out a second review to ascertain if concerns were addressed. Sixteen out of the 28 requests had been finally approved as of June 30, 2023. Most of the requests (17) received a conditional approval as first feedback, while 7 were asked for major revision.

- **Rejection** - a single DAR received an outright rejection.

From the requests that received a first decision of conditional approval or major revision, we examined the first decision letter sent by the DBAC secretariat based on the recommendations of DBAC's members' individual reviews and the main reasons for the decision were noted. Four major categories of reasons for only conditional approval were noted: (1) Research question and/or methodology needed clarification, (2) Ethics certificates were not provided or needed updating, (3) Benefits to Africa were not discussed by the applicant or were deemed insufficient or inadequate by the DBAC and, (4) Collaboration with African scientists was not evidenced. According to the DAR application form (section 6) the requester should justify the benefits to Africa and particularly how the proposed use of data and specimens could improve the health of the African population and contribute to the development of expertise and research capacity among African scientists. Appendix A of the guidelines to DBAC [9] also specifies that the DBAC should check that the requester has collaboration with an African scientist, although this is not considered a mandatory requirement.

Among the 25 requests that received the first feedback with conditional approval or revision, 9 (36%) were asked to provide a clearer description of the research question and/or the methodology for data use and analyses. Half (50%) of the requests required clarification around ethical issues, with requirements ranging from providing a valid ethical clearance certificate from their host institution, to clarifying risk of stigmatization or harm to African populations who provided the samples. Benefits to Africa were questioned in almost 40% of the DARs, while lack of clear African collaboration was raised in 61% of cases; these two issues often arose together (in 11 cases out of 25). All four issues were noted in 2/25 requests (and received a major revision decision). It is worth noting that the ethical issue comes as a single comment in 6 of the requests (Figure 3).

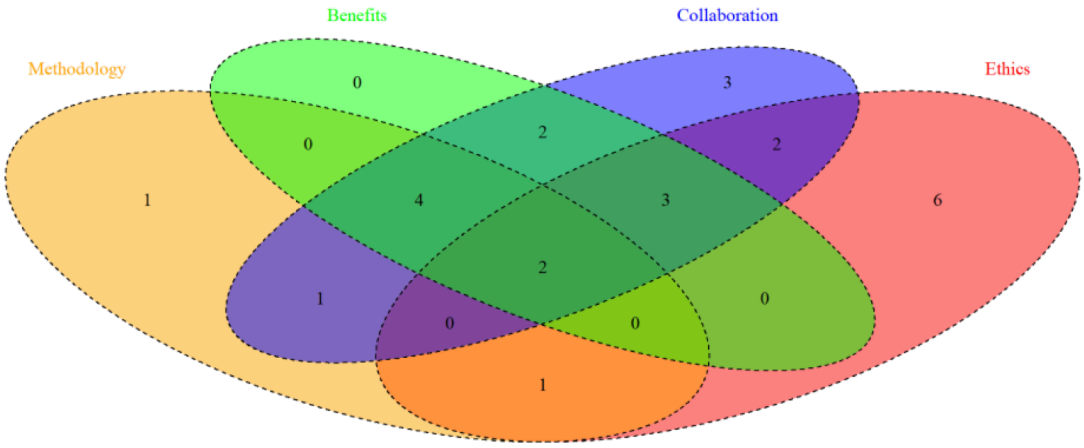

Figure 3. A Venn Diagram showing the categories of issues raised in the first feedback letter of DBAC to requesters

DAR from private institutions required clarification on "Methodology" significantly more frequently than DARs coming from the academic sector (71.4% versus 20%, respectively, Fisher-Exact Test (FET) p-value=0.023). The same trend is seen for the collaboration issue, with 100% of DARs from the private sector needing further input on this requirement, as opposed to 50% of DARs from the academic sector (FET p-value=0.026). Conversely, no significant difference was seen for the Ethics and the Benefits issues (FET p=0.67 and p=0.084, respectively).

Discussion

Origin of the requests

It is not surprising that most DARs came from the USA and UK, given that:

- The UK and the USA funded the H3Africa project.
- The two countries are active leaders in human genetics research.
- Most African PIs have a history of research collaboration with scientists in these two countries and that some of the work of H3A data generation (sequencing and genotyping) was done in collaboration with scientists from these countries [10].

DARs reflect an awareness of the H3A African data, and its unique ability to add value to a diverse range of research applications, particularly the establishment of allele frequencies in various normal populations and its potential utility in resolving the pathogenicity of variants. It is hoped that with increasing publications from the secondary use, this will further promote and advertise the existence of the African data beyond

only the USA and the UK. The few requests recently from countries such as Japan, India and Brazil might indicate that publications from the H3Africa consortia are being noticed by a wider readership, particularly those appearing in high impact journals [11][10][12]. The low number of requests from Africa itself might reflect lack of computational capacity and or expertise on the African continent. It is also possible that African scientists involved in data generation within the H3A consortium might have informally shared these data with their African collaborators directly.

Issues raised by DBAC about DARs

It is interesting to note that the nine DARs that received comments on three issues or more are equally distributed between the private and public sectors, indicating that the DBAC examined the DARs with similar criteria independently of their origin. However, it is significant to note that the methodology and the collaboration issue have been more frequently raised for DARs from the private sector. The methodology is likely explained by the fact that most DARs from private sector have not provided clear description of the research question and/or the data analyzes methods to be used, either intentionally so as not to reveal some of the tools they have developed, or the information they are looking for, or simply through lack of awareness of the importance of this issue.

The collaboration issue was raised for all DARs from the private sector and 50% of DARs from the academic sector. This reflects the DBAC's priority with respect to stimulating and ensuring collaboration with African scientists who produced these data and at least a minimal level of return of benefit to Africa. Although collaboration with African scientists is not a mandatory requirement to have access to data, the DBAC have given this issue high importance by categorizing collaboration with African scientists into six groups. If none of them exists, requesters are encouraged (in the feedback letter) to initiate collaboration. These six levels are: (1) new or already existing collaboration (officially funded or not) with one African scientist as a Co-PI or collaborator; if the collaboration is "non-official", a letter from the African scientist is acceptable evidence (2) common publications, (3) evidence of regular meeting or exchange, (4) training of African scientists within the requester's lab or institution, (5) lectureship or a visiting fellowship of the requester to an African institution and, (6) contribution to the infrastructure or capacity building in Africa.

It was noted that some of the requests seemed to overlook potential group harm depending on how data was intended to be used, and how findings were to be disseminated. Most requests seemed to overlook this issue, failing to appreciate the potential for group harm. Requesters were required to ensure responsible and sensitive dissemination in ways that did not stigmatize or spotlight certain groups. There is a history of exploitative research in Africa, and oversight bodies need to protect vulnerable populations [13][14]. At the same time, research efforts which benefit these same populations should not be hampered by overly

paternalistic committees, and equitable data sharing must be encouraged and facilitated. The WHO has set up the Technical Advisory Group on Genomics (TAG-G) [15] who, amongst other objectives, will shortly publish a document outlining guiding principles of human genome data sharing.

Conclusion

The H3Africa DBAC has played a role in ensuring that African genomes are responsibly interrogated for the benefit of the human population. The experience of this DBAC provides a template upon which new projects, or consortia, can build. In addition, the experience of this DBAC is now available to other upcoming initiatives and can provide advice or support, so that there is coordinated access to African data that are beneficial to the broader scientific community. It is hoped that increased availability of African genomics data, and its access to global researchers and companies, could benefit not only Africans, but humanity at large.

Statements and Declarations

Acknowledgments

The authors would like to thank Tumani Corrah and Alliance Nikuze as former members of the H3Africa DBAC for their contribution regarding the work that allowed establishment DBAC activities, and Christine Wasuna, Ayaga Bawah, Cornelius Debpuur and Phyliss Babirye as current members of the DBAC for their comments on the paper. We are also grateful to the H3Africa steering committee for the rich discussion, guidance, and interactions during and after the setup of the DBAC. The H3Africa DBAC has been funded by the H3Africa consortium and sustained by the Science for Africa Foundation.

Conflict of Interests

The authors declare no conflict of interest.

Authors' contribution

AR drafted the manuscript, analyzed the data, and produced table and figures. RW collected the data. KH produced figure 1. All authors conceptualized and edited the manuscript, revised it critically for important intellectual content and made substantial contribution to the concept or design of the article, or the acquisition, analysis, or interpretation of data for the article.

All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. ^*Chandramohan, D., Shibuya, K., Setel, P., Cairncross, S., Lopez, A. D., Murray, C. J., Zaba, B., Snow, R. W., & Binka, F. (2008). Should data from demographic surveillance systems be made more widely available to researchers? PLoS Medicine, 5(2), e57.*
2. ^*https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf*
3. ^*Shabani, M., Thorogood, A., Borry, P. (2016) Who should have access to genomic data and how should they be held accountable? Perspectives of Data Access Committee members and experts. Eur J Hum Genet 24, 1671–1675.*
4. ^*Cheah, P.Y., Piasecki, J. (2020) Data Access Committees. BMC Med Ethics 21, 12.*
5. ^*The H3Africa Consortium (2014) Enabling the genomic revolution in Africa. Science, 344: 1346-1348.*
6. ^*Wonkam A. (2021) Sequence three million genomes across Africa. Nature; 590: 209-211.*
7. ^*https://catalog.h3africa.org/*
8. ^*https://ega-archive.org/dacs/*
9. a, b, c*https://h3africa.org/wp-content/uploads/2020/06/App-D-H3Africa-Data-and-Biospecimen-Access-Committee-Guidelines-2020-.pdf*
10. a, b*Choudhury A., Aron S., Botigué L.R. et al. (2020) High-depth African genomes inform human migration and health. Nature. 586: 741 –748*
11. ^*https://h3africa.org/index.php/resource/publication-2/*
12. ^*Fortes-Lima, C.A., Burgarella, C., Hammarén, R. et al. (2024) The genetic legacy of the expansion of Bantu-speaking peoples in Africa. Nature. 625: 540-547.*
13. ^*Chennells R. and Steenkamp A. (2018). International Genomics Research Involving the San People, in D Schroeder, J Cook, F Hirsch, S Fenet and V Muthuswamy (eds) Ethics Dumping Case Studies from North- South Research Collaborations, Springer Briefs in Research and Innovation Governance, 99-106.*
14. ^*Keymanthri M. et al. (2022) Ethics and governance challenges related to genomic data sharing in southern Africa: the case of SARS-CoV-2. The Lancet Global Health, 12: e1855 - e1859*
15. ^*https://www.who.int/groups/technical-advisory-group-on-genomics-(tag-g)*

Declarations

Funding: The DBAC is funded by the Science For Africa Foundation.

Potential competing interests: No potential competing interests to declare.