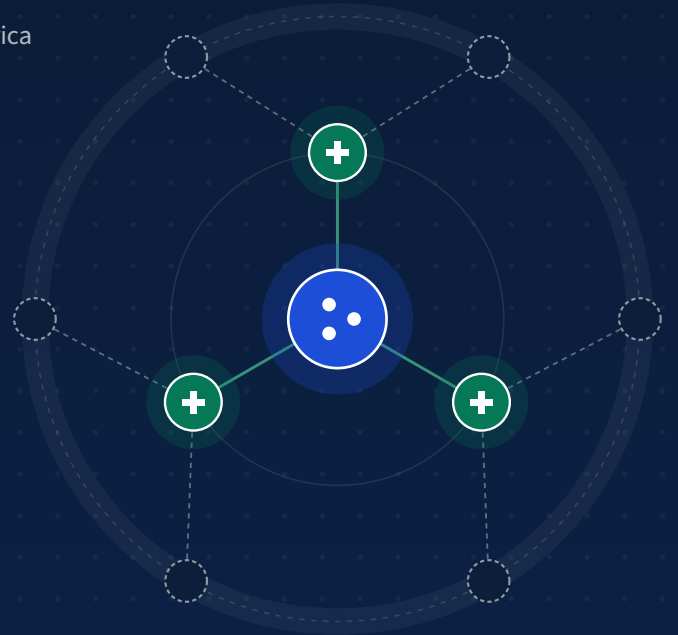


AfriBiobank

Medical imaging biobank for Africa



FUNDER BRIEF

● Coordinator ● Pilot sites, Phases 1 to 2 ○ Expansion, Phase 4

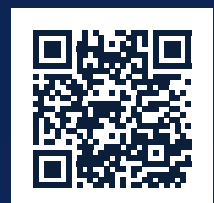
African imaging data. Kept **at home**. Working for science.

A peer-reviewed concept with a documented and independently assessed design for a governed, federated medical imaging infrastructure, in which hospitals keep their data and researchers still get answers. AfriBiobank now seeks initial funding to begin implementation and expand it across sites.

30 September 2026

Stage: design documented and independently reviewed; seeking initial implementation funding

Project website: <https://afribiobank.web.app>



Scan to visit
the project website

At a glance

Medical imaging such as X-ray, computed tomography (CT) and magnetic resonance imaging (MRI) sits at the centre of modern diagnosis, and it is a major source of training data for clinical artificial intelligence (AI). Large, well-governed imaging collections exist in Europe and North America. To the knowledge of the project authors, no imaging resource of comparable scale exists in Africa [1]. AfriBiobank is a documented plan to build one in a form that African institutions and regulators can accept: the images stay in the hospital and country where they were created, and only approved questions, summary answers and model updates travel.

The problem

African imaging data are scattered across institutions, rarely described in a common format and seldom shareable under current law and practice. As a result, clinical AI is mostly built and tested on patients from elsewhere, and the extent to which it performs well for African patients remains largely unmeasured.

The solution

A network of locally hosted nodes, one per participating institution or country, joined by common data standards, a shared catalogue and strict governance. Researchers ask questions of the network; each node answers under its own rules; raw images never leave.

Where it stands

Peer-reviewed concept [1], a detailed data model, draft system specifications, an independent technical review and a browser-based prototype running on synthetic data. No patient data have been collected and no pilot site has started.

Support sought

A 24-month programme to take the design to a working pilot: governance and ethics approvals, one reference hospital node, a two-to-three site federation, and a first federated AI study on a single, clinically important task.

1.5 billion

People in Africa in 2024 [2]

0

Raw images leaving a node in routine use (design goal)

24 months

From funding to a first multi-site federated study

6 countries

Collaborators in 4, discussions in 2

Why fund this now

The main technical ingredients (open imaging standards, federated learning software and privacy engineering methods) are available and have been used in medical research [11], and the core design has been drafted and independently reviewed. What is missing is the funded effort to secure approvals, stand up real sites and demonstrate, on real data and under real governance, that the approach works. That step is where most of the remaining uncertainty lies, and it is the step this programme would fund.

1 The problem

1.1 Imaging and AI depend on representative data

Imaging studies are among the most information-rich records in medicine. They support diagnosis, treatment planning and follow-up, and they are among the most common inputs to clinical AI. AI models learn from the examples they are shown. When those examples come mainly from one population, one set of scanners and one set of clinical habits, the model's performance elsewhere is uncertain until it has been tested.

There is evidence that this concern is justified. In a widely cited study, chest X-ray classifiers trained on large public datasets were more likely to label patients from under-served groups as healthy when they were not, including Black patients, female patients and patients with public insurance [3]. That work used North American data. Comparable evidence for African patients appears scarce, in part because the data needed to produce it are not accessible. This is a central part of the problem AfriBiobank addresses.

1.2 A missing resource

Open and controlled-access imaging resources such as UK Biobank [17], the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Brain Tumor Segmentation benchmark (BraTS) have supported a large body of imaging research [4]. Africa has important biobanking efforts, notably within the Human Heredity and Health in Africa (H3Africa) initiative. The Stroke Investigative Research and Educational Network (SIREN), for example, holds more than 3,000 brain images collected in Nigeria and Ghana [5]. These efforts were designed mainly around biological samples and genetic data, and the AfriBiobank authors describe them as largely static databases with limits on scalability, security and high-volume access [1]. To the authors' knowledge, a continent-wide, governed resource built for imaging does not yet exist.

1.3 Why the data do not move

The AfriBiobank analysis, first outlined in 2023 [23] and later published in peer-reviewed form, identifies six interacting barriers [1, 6, 7]. They are summarised in Figure 1. Most of them are not purely technical, so software alone is unlikely to be sufficient.

Equipment is a further constraint. A continent-wide assessment of MRI capacity reported gaps in training, maintenance and research capacity across sub-Saharan Africa [8]. Any imaging infrastructure therefore has to cope with varied and sometimes older equipment, and with images of uneven quality.

1.4 What this costs

Without accessible local data, African researchers compete for scarce public datasets from elsewhere, AI tools are adopted with little local validation, and diseases that are common in Africa but rare in high-income settings receive limited attention in imaging research. Imaging data are generated every day in African hospitals. The gap therefore seems to lie less in the existence of



Figure 1. Six barriers to sharing medical imaging data in Africa, as identified in the literature reviewed by the project [1, 6, 7].

data than in the absence of a trusted, lawful way to use them together.

2 What AfriBiobank is

Vision

A future where every diagnostic AI system understands African patients as well as any other population, so that data support more equitable healthcare outcomes across the continent.

Mission

To widen access to African medical imaging data while preserving privacy and data sovereignty, so that researchers and AI developers can build solutions that work for African populations.

Values. Equity and inclusion; privacy and sovereignty, in the sense that data belong to the institutions and communities that generate them; collaboration across borders; innovation grounded in open science; ethical leadership with community oversight; and capacity building rather than dependency (project website, afribiobank.web.app/about).

2.1 The central idea: data stay, knowledge travels

Most data-sharing models copy data into one central repository. For health data in Africa this raises serious legal, ethical and practical concerns, because it asks institutions and countries to give up physical control of their patients' records. AfriBiobank reverses the direction of travel (Figure 2). Each participating institution runs its own node. Questions and model-training tasks are sent to the nodes. Each node decides, under its own rules, what it will answer, and returns only what is permitted: descriptions, counts that are large enough to be safe, or updates to a shared AI model.

2.2 Four building blocks

The design has four parts that work together (Figure 3).

Governance and ethics. The rules that decide who may ask what, for which purpose, under which consent and agreement. They include a data access committee, consent records and benefit-

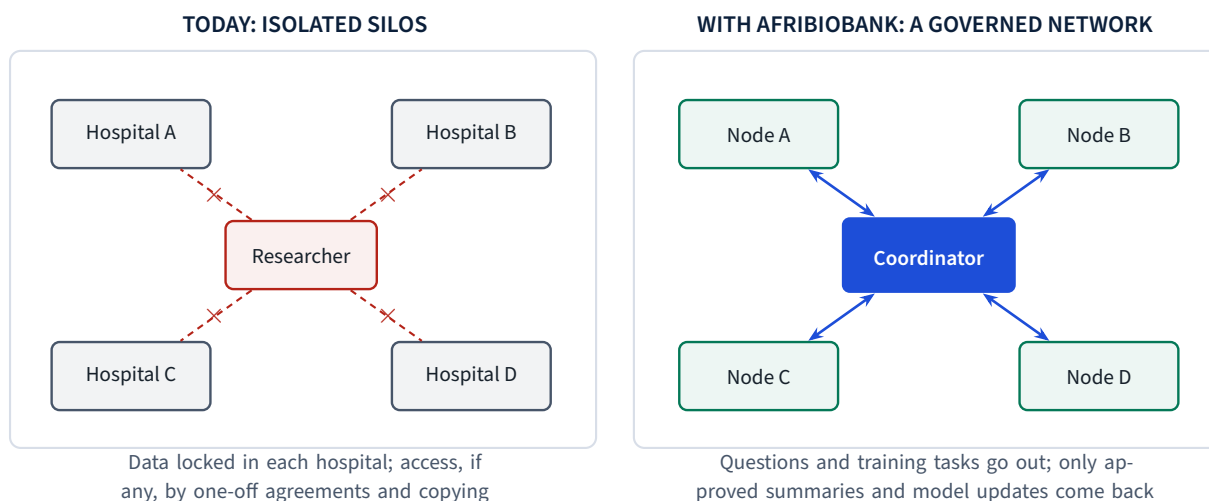


Figure 2. From isolated silos to a governed network. On the right, each node keeps its images and applies its own access rules before answering.

sharing terms. Governance applies to every other layer.

Local node. Hardware and software hosted at the institution or within the country, holding the images and their descriptions. A node is sized to the site, from a single server for a pilot hospital to a small cluster for a national centre.

Semantic catalogue. A common way of describing patients, studies, images and findings so that data from different hospitals can be searched together. It relies on international standards: Digital Imaging and Communications in Medicine (DICOM) for images, Fast Healthcare Interoperability Resources (FHIR) for clinical records, and the Resource Description Framework (RDF), a standard format for linked, machine-readable data [1, 9].

Federated analysis and learning. The means by which questions and AI training reach the nodes and results come back, without raw data moving (Section 2.4).

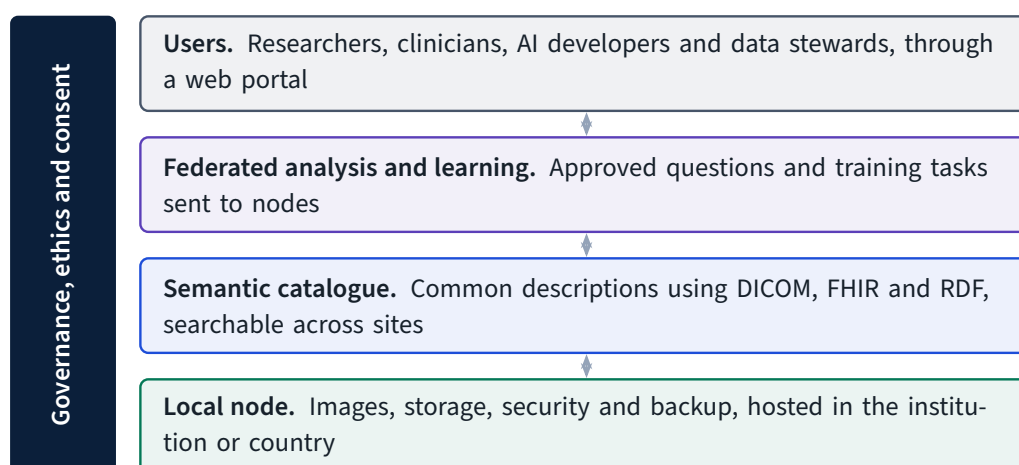


Figure 3. AfriBiobank as four layers. Governance is drawn as a spine because it constrains every layer, not only the user-facing one.

2.3 How a researcher would use it

Figure 4 shows the intended path for a research question. The same path applies whether the requester is at an African university, a hospital or an international partner.



Figure 4. The intended researcher journey. Review happens before any data are touched, and every step leaves an audit record.

2.4 Federated learning in plain terms

Federated learning is a way of training one AI model across several institutions without pooling their data [10, 11]. A coordinator sends the current model to each site. Each site trains it for a short time on its own images and sends back only the resulting changes to the model, not the images. The coordinator combines these changes, weighting each site by how much data it contributed, and the cycle repeats (Figure 5). In a brain tumour segmentation study across ten institutions, a federated model was reported to reach about 99% of the performance of a model trained on the same data pooled centrally [12]. Results of this kind are encouraging, although performance in any new setting has to be measured rather than assumed.

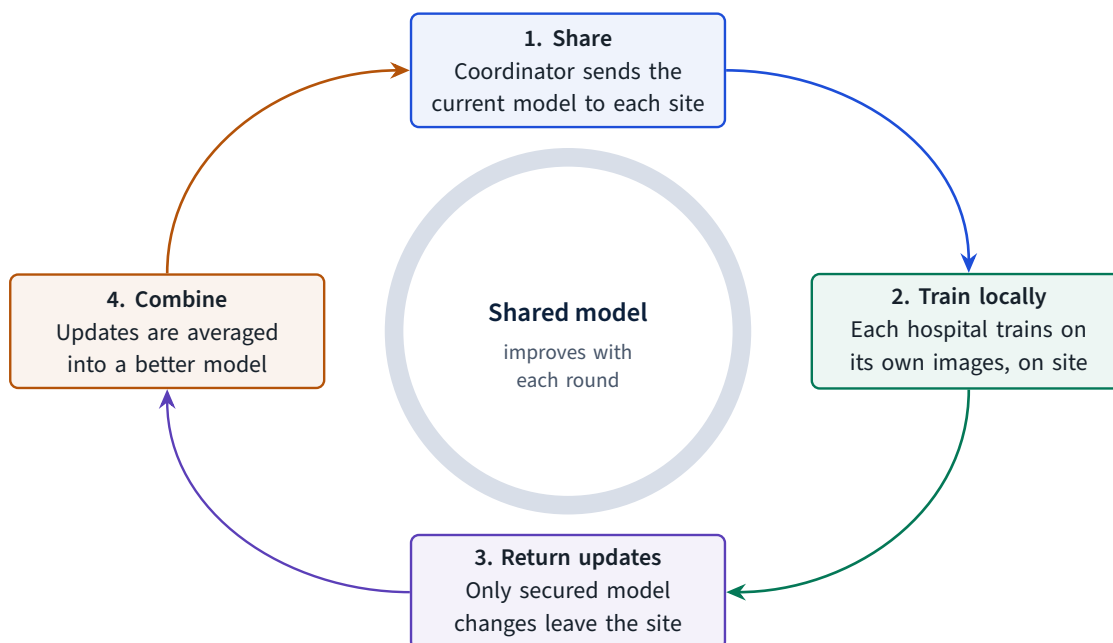


Figure 5. One round of federated learning. Images stay at each site throughout; only model parameters travel, and these are protected in transit.

Federated learning is not, by itself, a complete privacy guarantee. Research has shown that model updates can in some conditions leak information about the training data [13]. The design therefore treats additional protections, such as secure aggregation and differential privacy

(a mathematical limit on how much any one patient can influence a result [14]), as measured design choices rather than assumptions (Section 3).

2.5 Learning from established models

AfriBiobank draws on three kinds of working precedent (Table 1). Each shows that part of the design is feasible; none combines all of the elements that African settings require.

- **UK Biobank** holds health data on about half a million participants and, since 2014, has been re-inviting 100,000 of them for brain, cardiac and abdominal imaging; more than 45,000 had been imaged by early 2020 [16, 17]. It shows the scientific value of a large, well-governed imaging resource, but it is a centralised national collection.
- **OpenSAFELY**, run by the Bennett Institute at the University of Oxford with NHS England and technology partners, analyses NHS records inside a secure environment: researchers send code to the data rather than receiving the data [18, 19]. It reports more than 200 projects and more than 100 published outputs, which suggests that this “code to data” model can work at national scale.
- **Flywheel** is a commercial platform for managing medical imaging research data, offering cohort discovery, curation and secure sharing across sites [20]. It shows that imaging data tooling is mature, but as a licensed, cloud-based product its cost and hosting model may be difficult to sustain in many African institutions.

Feature	UK Biobank	OpenSAFELY	Flywheel	AfriBiobank (planned)
Imaging focus	yes	no	yes	yes
Data stay where they were collected	no	yes	depends on deployment	yes
Analysis sent to the data	partly	yes	depends on deployment	yes
Open-source core	no	yes	no	yes
Designed for African law and infrastructure	no	no	no	yes
Operating today	yes	yes	yes	no

Table 1. Comparison with established models, based on their published descriptions. “Partly” marks a feature available for some uses but not by default; “depends on deployment” marks a feature that varies with how the product is installed. The last row is included deliberately: AfriBiobank has yet to demonstrate what the others already have.

2.6 Grounding in the published work

The design builds on the peer-reviewed AfriBiobank paper [1], which made three contributions: an analysis of the barriers to imaging data availability in Africa, a multi-stakeholder biobank model that joins technical development with governance, and an outline of the software and hardware needed to implement it. Its policy recommendations, which this programme would

put into practice, include national imaging biobanks under common ethical oversight, broad informed consent in the participant's own language, training in data science, and harmonised data regulation within regional economic zones. The paper is exploratory: it describes the design and defers a pilot to follow-up work. The programme in Section 7 is that follow-up.

3 Protecting patients and communities

3.1 Three levels of access

Not every question needs the same level of detail. The design defines three classes of access, each with stronger controls than the last (Figure 6). Many research questions, such as “how many chest X-rays with a tuberculosis finding exist across the network, by age group”, can be answered at the aggregate level without any individual record leaving a site.

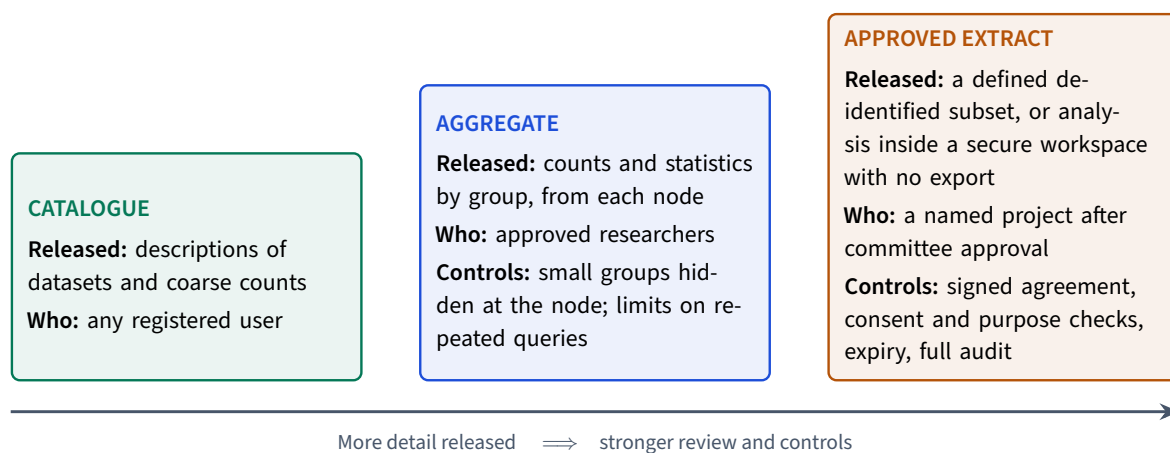


Figure 6. Three access classes. Detail increases from left to right, and so do the review and controls attached to it.

3.2 De-identification of images

Removing names from a file is not enough for medical images. Patient details can be printed into the pixels of an X-ray, and a head CT or MRI scan can be rendered into a recognisable face. In one study, face-recognition software matched MRI-based reconstructions to photographs for 70 of 84 research participants [15]. Before any data are loaded, the design therefore applies the international DICOM confidentiality profile (PS3.15), detects text burned into image pixels, removes facial features from head scans, shifts dates consistently for each patient, replaces identifiers with coded pseudonyms whose key never leaves the site, and prevents very small groups from being reported. Residual risk will be assessed and documented for each type of release. No method removes risk entirely, so the aim is to reduce it to a level that ethics committees and data protection authorities judge acceptable.

3.3 Consent, law and benefit sharing

Consent and permitted use are recorded with the data, using the FHIR Consent standard and the Data Use Ontology (DUO) of the Global Alliance for Genomics and Health (GA4GH), which expresses terms such as “general research use” in machine-readable form [27]. Following the published recommendation, consent forms will be available in participants’ own languages, or explained with a translator, so that agreement is informed [1]. A patient’s withdrawal of consent removes their data from future use and is recorded in the audit log.

Legal obligations differ by country, and the legal landscape is changing quickly. The Data Protection Africa tracker, maintained by ALT Advisory, listed 43 data protection laws across 55 African countries in January 2026 [21], and the non-profit Africa Data Protection monitors legislation in 54 countries [22]. Relevant instruments include, for example, South Africa’s Protection of Personal Information Act (POPIA), Nigeria’s Data Protection Act 2023, Kenya’s Data Protection Act 2019 and Ghana’s Data Protection Act 2012, together with the African Union Data Policy Framework. The programme budgets for legal review in each pilot country, since these laws change and their interpretation for research varies.

The published governance model assigns defined roles to not-for-profit organisations, government ministries and agencies, and contributing hospitals [1]. Terms for benefit sharing, including how contributing sites are credited and whether any commercial access is permitted, will be agreed with partners before such access is built. They are not assumed in advance.

4

Who benefits

AfriBiobank is multi-sectoral by design [1]. Primary stakeholders shape policy, create and hold data, and run the infrastructure. Secondary stakeholders use its outputs or benefit from them indirectly (Figure 7).

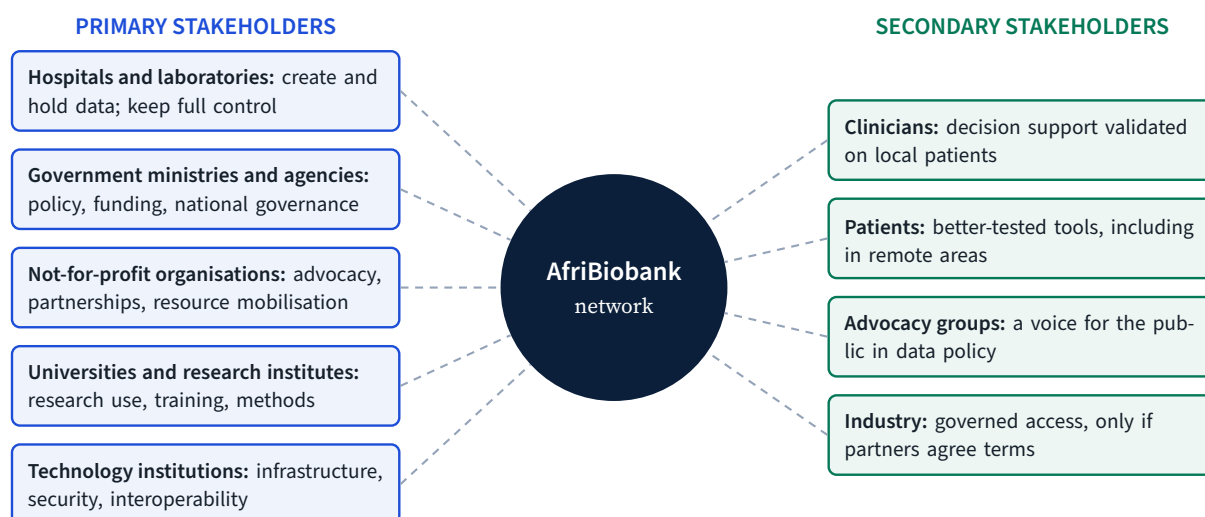


Figure 7. Stakeholders, grouped as in the published governance model [1]. Industry access is shown as conditional because its terms have not yet been agreed.

5 Where the project stands

AfriBiobank's design is documented and independently reviewed, and the project is ready to begin implementation. The foundations below are in place; the pilot that would test them on real data has not started. An accurate baseline is the fairest basis for a funding decision, so both lists are given below. AfriBiobank began during Dr Lukman Enegi Ismaila's fellowship at the Datasphere Initiative [25]. An article published by the Initiative on 25 January 2023 analysed the barriers to medical data for science in Africa [23], and a second, on 3 July 2023, set out the idea of a multi-actor medical imaging biobank [24], and the later peer-reviewed paper acknowledges the Initiative's support [1]. The concept was first presented at the MICCAI Meets Africa workshop (MICCAI is the Medical Image Computing and Computer Assisted Intervention conference) and later reached the final three at MIRASOL 2025, the NextGen Interchange on Medical Image Computing in Resource-Constrained Settings.

Collaboration footprint

AfriBiobank has active collaborators in Nigeria, Ghana, South Africa and Tunisia, and discussions are under way with partners in Kenya and Uganda (Figure 8). Pilot sites will be chosen from among these institutions during Phase 0. Collaboration does not yet mean that data-sharing agreements have been signed; those agreements are a Phase 0 deliverable.

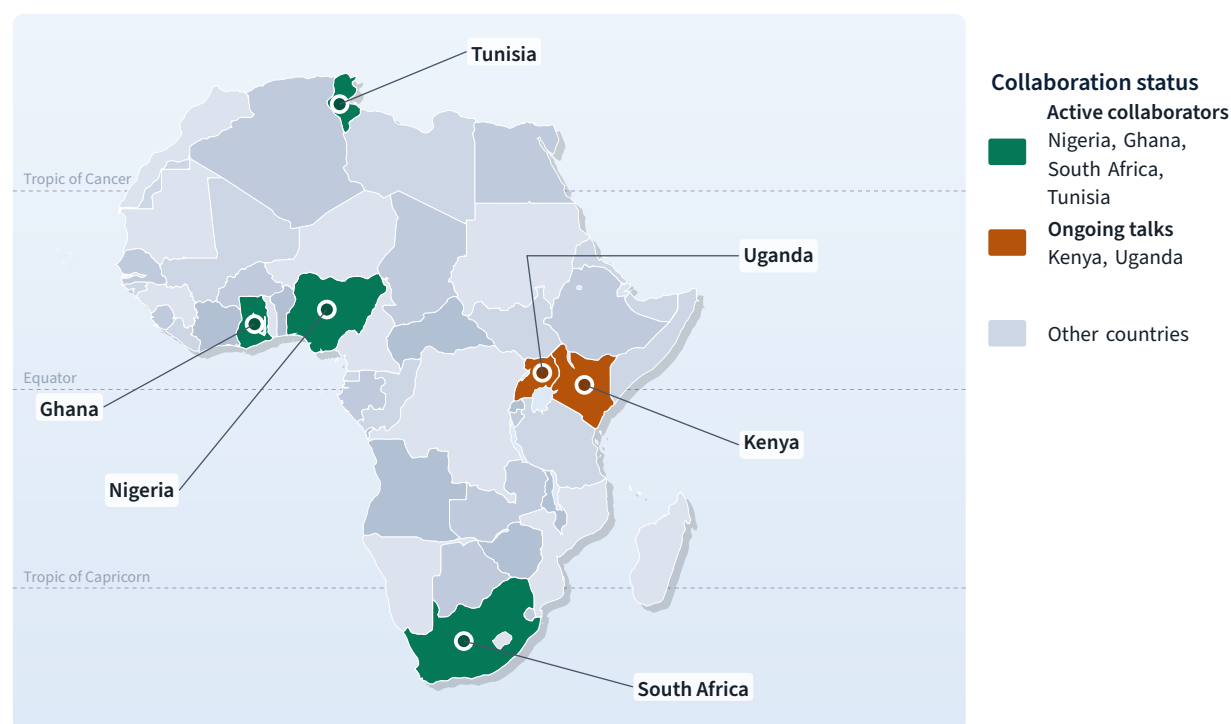


Figure 8. Current AfriBiobank collaborations. Country boundaries from Natural Earth (1:50 million scale), equirectangular projection. The map shows where collaborators are based; it does not indicate that data have been shared or that agreements are in place.

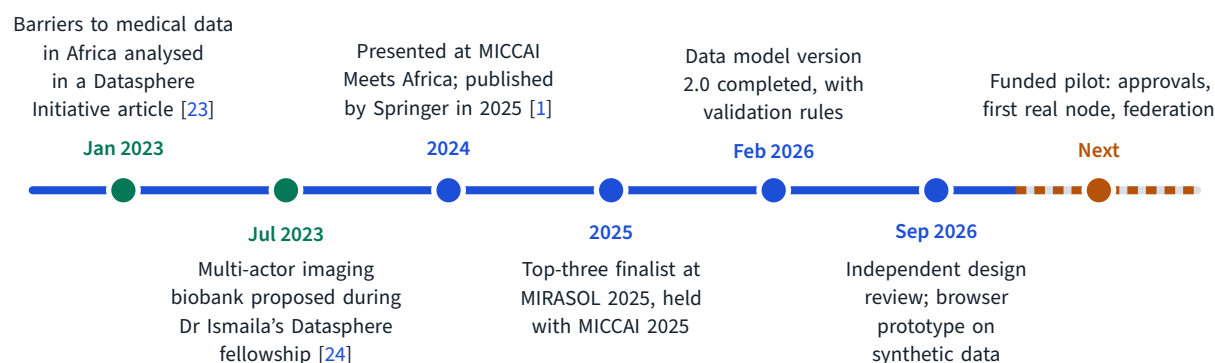


Figure 9. Progress to date, from the first analysis of the problem in January 2023 to the present. The dashed segment is the implementation this funding would start.

In place today

- A peer-reviewed concept and governance model [1].
- A detailed data model (version 2.0) describing patients, studies, images, findings and their provenance, with automated validation rules.
- Draft specifications for the software, federation and deployment.
- An independent review (September 2026) that identified 29 issues in the drafts, eight of them rated high severity, and proposed fixes before any build began.
- A working browser prototype on synthetic data that demonstrates federated counting with small-group suppression, role-based access and a tamper-evident audit log.
- A public website with a demonstration dashboard that runs on clearly labelled sample data.

Not yet in place

- No patient data have been collected or loaded.
- Pilot sites among the collaborating institutions are still to be confirmed, and no data-sharing agreement has been signed.
- Ethics approvals and country-level legal reviews have not yet been sought.
- The production software (node, gateway, federation service) has not been built.
- No AI model has been trained within AfriBiobank.

These are the gaps the proposed programme is designed to close.

The independent review is itself a strength of the project. It found, for example, that early drafts relied on two competing databases, placed access control where it could be bypassed and had not yet specified image de-identification. Each issue now has a documented remedy, and the plan in Section 7 is built on the corrected design. Finding such problems at the design stage is considerably cheaper than finding them after patient data have been loaded.

6 The case for investment

Six arguments, each grounded in the evidence above, support funding AfriBiobank at this stage.

1. A documented gap

To the authors' knowledge, Africa has no governed imaging resource comparable to UK Biobank, ADNI or BraTS [1], while imaging AI has been shown to under-serve groups that are poorly represented in its training data [3].

2. A large and growing population

Africa had about 1.5 billion people in 2024, and United Nations projections put sub-Saharan Africa alone at about 2.2 billion by 2054 [2]. Clinical AI that has not been tested on this population is unlikely to serve it reliably.

3. Already de-risked

The concept is peer-reviewed [1] and was a top-three finalist at MIRASOL 2025. The data model is on its second version, 29 design issues were found and resolved on paper before any build, and a working prototype demonstrates the core privacy controls.

4. Built for sovereignty

Because images never leave the hospital or country, the design works within data localisation and cross-border transfer rules rather than against them, which lowers a barrier that has stalled many sharing efforts [6].

5. Investment that compounds

Each node, toolkit and agreement template is reusable. The first site carries most of the design cost; later sites reuse the node toolkit and onboarding kit, and other African networks can adopt the open-source components.

6. Clear accountability

Every phase has exit criteria, and progress is tracked against the key performance indicators (KPIs) in Section 9.1, each with a baseline, a target and a means of verification. Funders can see whether the programme is working at each gate.

7 The plan

7.1 Principles

The plan builds one real site end to end before adding others, lets governance gate the use of real data (software is built on synthetic data meanwhile), treats privacy as something to be tested rather than asserted, and sizes nodes so that a modest hospital can run them. It starts with one clinical task: a candidate is chest X-ray triage for tuberculosis or pneumonia, because chest radiographs are common, relatively small files and clinically important in many African settings. The final choice will be made with pilot sites.

7.2 Phases and exit criteria

The programme runs over 24 months in five overlapping phases. Durations depend on team size and on how quickly pilot-site agreements are signed, which is the main source of schedule uncertainty. Each phase ends only when its exit criteria are met, so that funders can see progress against evidence rather than activity.

Phase	Months	Main deliverables	Exit criteria
0 Foundations	0 to 2	Governance charter, data access committee, ethics route, legal review started, pilot sites selected, design decisions recorded	Agreement template approved; pilot sites committed
1 Reference node	2 to 6	One real hospital end to end: image archive, de-identification, catalogue, access gateway, viewer, audit log	Studies flow from the hospital archive to the catalogue at a stated pass rate; privacy attack tests refused
2 Federation pilot	6 to 12	Two to three sites connected; catalogue and aggregate queries; consent filters; committee workflow; external security test	A cross-site question answered on real data under approval; no critical security findings open
3 Federated learning	9 to 15	One clinical task trained across sites; per-site and subgroup evaluation; model documentation	Federated model at least as good as local models at held-out sites; no raw data moved
4 Scale and sustain	12 to 24	More sites and countries, researcher portal, training programme, sustainability plan, publications	Operating model and funding agreed for years 3 to 5; a new site onboarded using the playbook alone

8 What funding would enable

The programme is organised into five work packages. A costed budget will be prepared for each specific call, because hardware, hosting, personnel and legal costs differ markedly between countries and would be estimated with the confirmed pilot sites.

WP	Work package	What it delivers
WP1	Governance, ethics and law	Data access committee; consent and data-use terms; country legal reviews; data-sharing and benefit-sharing agreements
WP2	Reference node	Hardware and software for the first hospital node; de-identification pipeline; security testing
WP3	Federation	Secure connection of two to three sites; aggregate queries with small-group protection; audit
WP4	Federated AI study	One clinical task, evaluated per site and per patient subgroup, with documented limitations
WP5	Capacity and sustainability	Training for site teams; onboarding playbook; open-source release; long-term operating and cost-sharing model

Planned toolkits

Alongside the network, the programme would produce six reusable toolkits, each released under an open licence where partners agree. None of them exists yet as working software. Phase 1 would deliver a **node toolkit** (archive, catalogue, access gateway, audit log and backup, sized from one server upward), an **imaging preprocessing and de-identification pipeline** and a **quality control toolkit**. Phase 2 would add a **site onboarding and mapping kit**, Phase 3 a **federated learning toolkit** with per-site and subgroup evaluation, and Phases 3 to 4 documented **AI-ready outputs**: curated cohorts, labels and, where approved, segmentations.

Potential AI tools

The project website outlines a portfolio of AI tools that the network could support once governed data are available [26]. They are potential applications, not validated products: each would need training and evaluation on African data, under the governance in Section 3, before any clinical use. The first to be pursued will be the single task chosen for Phase 3.

Clinical imaging applications

- **Chest X-ray analysis:** tuberculosis, pneumonia and other thoracic findings
- **Brain tumour detection:** detection, classification and segmentation on MRI
- **Retinal disease screening:** diabetic retinopathy, glaucoma and macular degeneration
- **Skin lesion classification:** trained across diverse African skin tones
- **Cardiac imaging:** cardiac MRI and echocardiography analysis
- **Rare disease observatory:** monitoring of rare conditions across African populations

Enabling and research tools

- **Image quality assessment** before AI analysis or clinical review
- **Multi-organ segmentation** on CT and MRI
- **Disease progression prediction** from repeated imaging over time
- **Explainable AI:** visual explanations to support clinical decisions
- **Synthetic image generation** for training and privacy protection
- **Model deployment and inference services**, and **clinical trial matching**

9

Expected outcomes and key performance indicators

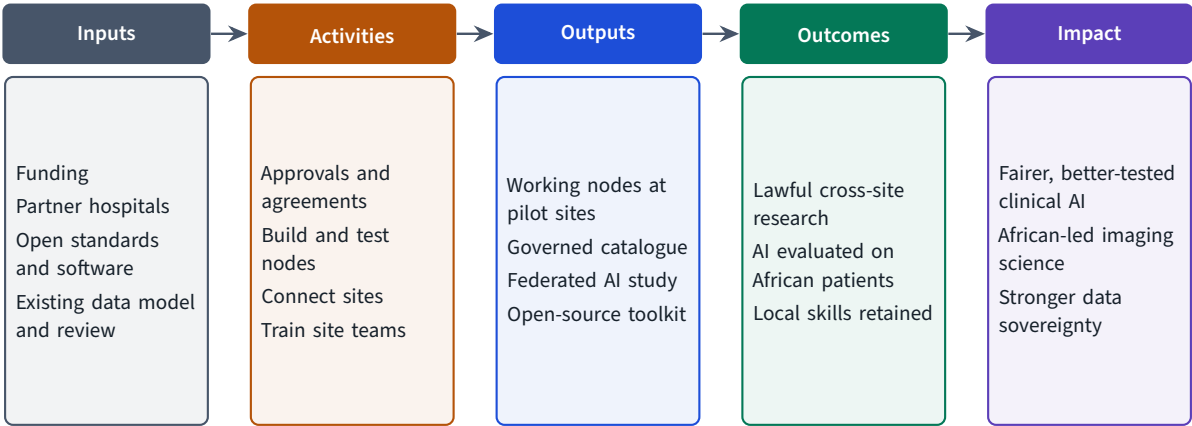


Figure 10. Logic model. Outputs are within the programme’s control; outcomes and impact depend also on partners, policy and time, and are therefore stated as intended rather than guaranteed.

9.1 Key performance indicators

Table 2 sets out the indicators the programme would report against. Baselines describe the position in September 2026. Targets are indicative: they follow from the plan in Section 7 and will be confirmed with funders and pilot sites during Phase 0. Volume targets, such as the number of imaging studies, will be set at the end of Phase 1, once the actual volumes of the pilot sites are known.

The ask in one sentence

Support for a 24-month programme that turns a peer-reviewed, independently reviewed design into a working, governed, multi-site imaging network, and that produces the first federated evaluation of clinical AI on African imaging data within that network.

10

Team and contact

AfriBiobank is led by Dr Lukman Enegi Ismaila. Active collaborators at the SisonkeBiotik research community (South Africa), the University of Sfax (Tunisia) and Nile University of Nigeria have made this work possible [1]. More active collaborators are showing interest in joining.

We thank [Houcemeddine Turki](#) and [Chris Fourie](#) for their effort and contribution to the possible pilot implementation plan.

ID	Indicator	Baseline	Month 12	Month 24	Verified by
Governance and trust					
K1	Data-sharing agreements signed with pilot sites	0	2	3 or more	Signed agreements
K2	Countries with completed data protection legal review	0	2	3 or more	Legal opinions
K3	Ethics approvals in place for data use	0	2 sites	All active sites	Approval letters
K4	Releases with an approved re-identification risk assessment	n/a	100%	100%	Committee minutes
Infrastructure and data					
K5	Nodes live on real, de-identified data	0	2	3 to 4	Node registry; audit log
K6	De-identified imaging studies in the governed catalogue	0	Set at month 6	Set at month 6	Catalogue counts
K7	Records discarded without a logged reason	n/a	0	0	Provenance records
K8	Quarterly backup restore drills passed	n/a	100%	100%	Drill reports
Privacy and security					
K9	Privacy attack queries refused or suppressed	Prototype only	100%	100%	Automated test suite
K10	Critical findings open after external penetration test	n/a	0	0	Test report
Research and AI					
K11	Time to answer a cross-site aggregate question	n/a	About 1 minute or less	About 1 minute or less	System logs
K12	Approved research projects using the network	0	1	3 or more	Committee register
K13	Federated model performance against local models at held-out sites	n/a	Study designed	At least equal at every site; subgroup gaps reported	Evaluation report
K14	Peer-reviewed papers or preprints reporting results	0	1	3 or more	Publication list
Capacity and openness					
K15	Site staff trained in imaging informatics and data governance	0	10 or more	25 or more	Training records
K16	Time to onboard a new site	n/a	Measured	8 weeks or less	Onboarding log
K17	Core toolkits released under open licences (of six)	0	2	6	Public repositories

Table 2. Key performance indicators. Baselines as of September 2026; targets are indicative and will be confirmed in Phase 0. “n/a” marks indicators that cannot be measured until the relevant component exists.

Contact: Dr Lukman Enegi Ismaila, lismail1@jhu.edu

Project website: <https://afribiobank.web.app>

A companion *Technical Document* with architecture, data model, privacy engineering and delivery detail is available on request.

References

- [1] Ismaila LE, Turki H, Frikha M, Weinstein T, Hunja F, Fourie C, Adeshina SA. AfriBiobank: empowering Africa's medical imaging research and practice through data sharing and governance. In: *MIImA 2024/EMERGE 2024*, Communications in Computer and Information Science, vol. 2240, pp. 189–198. Springer, 2025. doi:10.1007/978-3-031-79103-1_20
- [2] United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2024*. New York: United Nations, 2024.
- [3] Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine* 27:2176–2182, 2021.
- [4] Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34(10):1993–2024, 2015.
- [5] Akinyemi RO, Akinwande K, Diala S, et al. Biobanking in a challenging African environment: unique experience from the SIREN project. *Biopreservation and Biobanking* 16(3):217–232, 2018.
- [6] Staunton C, Moodley K. Challenges in biobank governance in Sub-Saharan Africa. *BMC Medical Ethics* 14:35, 2013.
- [7] Adebamowo CA, Callier S, Akintola S, et al. The promise of data science for health research in Africa. *Nature Communications* 14, 2023.
- [8] Anazodo UC, Ng JJ, Ehiogu B, et al. A framework for advancing sustainable magnetic resonance imaging access in Africa. *NMR in Biomedicine* 36(3):e4846, 2023.
- [9] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018, 2016.
- [10] McMahan B, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of AISTATS*, PMLR 54:1273–1282, 2017.
- [11] Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *npj Digital Medicine* 3:119, 2020.
- [12] Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific Reports* 10:12598, 2020.
- [13] Zhu L, Liu Z, Han S. Deep leakage from gradients. In: *Advances in Neural Information Processing Systems* 32, 2019.
- [14] Dwork C, Roth A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9(3–4):211–407, 2014.
- [15] Schwarz CG, Kremers WK, Therneau TM, et al. Identification of anonymous MRI research participants with face-recognition software. *New England Journal of Medicine* 381:1684–1686, 2019.
- [16] Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* 12(3):e1001779, 2015.
- [17] Littlejohns TJ, Holliday J, Gibson LM, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications* 11:2624, 2020.
- [18] Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584:430–436, 2020.
- [19] Bennett Institute for Applied Data Science, University of Oxford. OpenSAFELY. <https://www.opensafely.org/> (accessed 30 September 2026).
- [20] Flywheel. Medical imaging data management platform. <https://flywheel.io/> (accessed 30 September 2026).
- [21] ALT Advisory. Data Protection Africa. <https://dataprotection.africa/> (last updated January 2026; accessed 30 September 2026).
- [22] Africa Data Protection. <https://www.africadataprotection.org/> (accessed 30 September 2026).
- [23] Ismaila LE. State of medical data in Africa: a case of barriers to science in Africa. Datasphere Initiative, 25 January 2023. <https://www.thedatasphere.org/news/state-of-medical-data-in-africa-a-case-of-barriers-to-science-in-africa/>
- [24] Ismaila LE. Breaking barriers: empowering Africa's health research through data sharing and governance. Datasphere Initiative, 3 July 2023. <https://www.thedatasphere.org/news/breaking-barriers-empowering-africas-health-research-through-data-sharing-and-governance/>
- [25] Datasphere Initiative. Fellowship programme: Lukman Enegi Ismaila. <https://www.thedatasphere.org/fellowship-program/lukman-ene-gi-ismaila/> (accessed 30 September 2026).
- [26] AfriBiobank. AI tools. <https://afribiobank.web.app/ai-tools/> (accessed 30 September 2026).
- [27] Lawson J, Cabili MN, Kerry G, et al. The Data Use Ontology to streamline responsible access to human biomedical datasets. *Cell Genomics* 1(2):100028, 2021.