

Data Science for Health Discovery and Innovation in Africa (DS-I Africa)

Building infrastructure & capacity in Africa to support health systems

PMAC

1 February 2025

Michelle Skelton



National Institutes of Health

Office of Strategic Coordination – The Common Fund

\$88 M over FY21-25 | 38 awards across 22 African countries + U.S.



Research

- 7 Research Hubs
- 13 Partnership for Innovation Projects



Capacity Building

- 7 Research Training Programs
- 6 Education Programs



ELSI

- 4 Ethical, Legal, and Social Implications Research Projects



Ecosystem

- Coordinating Center
- eLwazi Open Data Science Platform

Research Hubs

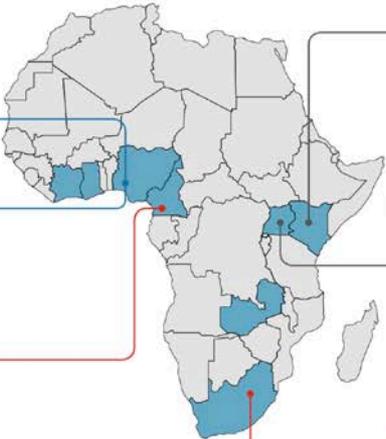
RESEARCH HUBS



Cameroon
Cote d'Ivoire
Ghana
Kenya

Nigeria
Rwanda
South Africa
Uganda

*Project has partner institution(s) in the United States of America



CAMRA* | Grant Number: 5US4TW012056-03
 Combating AntiMicrobial Resistance in Africa Using Data Science
Data Types: Demographic, Clinical/Phenotype/Climate/Genomic
Platform: elwazi, NCBI
Capacity Building: Training on Bacterial Sequencing & Bioinformatics
Focus Areas: Antimicrobial Resistance in Invasive Bacteria
Redeemer's University
 Nigeria

Contact PI: Christian Happi

INFORM Africa* | Grant Number: 5US4TW012041-03
 Role of Data Streams in Informing Infection Dynamics in Africa
Data Types: Demographic, Surveillance, Clinical/Phenotype, Environmental, Climate, Genomic, Image
Platform: Dropbox
Capacity Building: Research Ethics, Research/Bioinformatics Methods, Science Communication
Focus Areas: HIV, COVID-19, Population-Scale Data Streams
Institute of Human Virology - Nigeria

Contact PI: Alash'le Abimiku

D-SINE* | Grant Number: 5US4TW012087-03
 Harnessing Data Science to Promote Equity in Injury and Surgery for Africa
Data Types: Demographic, Surveillance, Clinical/Phenotype, Genomic, Context of Injury
Platform: HPC
Capacity Building:
Focus Areas: Injury, Trauma, Surgery, Health Equity
University of Buea - Cameroon

Contact PI: Melire Chichom

UZIMA-DS* | Grant Number: 5US4TW012069-03
 Utilizing health information for meaningful impact in East Africa through Data Science
Data Types: Demographic, Surveillance, Clinical/Phenotype, Environmental, Geospatial
Platform: Azure
Capacity Building: Azure Training, Data Journalism, Scientific Writing, Research Management Foundations
Focus Areas: Maternal/Newborn/Child Mental Health
Agri Khan University - Kenya

Contact PI: Amina Abubakar

MUDSReH* | Grant Number: 5US4TW012043-03
 MUST Data Science Research Hub
Data Types: Demographic, Surveillance, Clinical, Environmental, Image, Geospatial
Platform: Google Cloud Platform
Capacity Building: Implementation Science, Responsible Conduct of Research, Scientific Manuscript Writing, Grant Writing
Focus Areas: DCPA, Breast Cancer
Mbarara University of Science and Technology
 Uganda

Contact PI: Fred Kaggwa

MADIVA* | Grant Number: 5US4TW012077-03
 MADIVA (Multimorbidity in Africa: Digital innovation, visualisation and application)
Data Types: Demographic, Surveillance, Clinical/Phenotype, Genomic
Platform: WITS Cluster
Capacity Building: Scientific Writing, Grant Proposal Writing, Research Governance, Qualitative Data Analysis, Community and Stakeholder Engagement, Ethical and Regulatory Issues
Focus Areas: Multimorbidity, Polygenic Risk Scores
WITS Health Consortium (PTY), LTD
 South Africa

Contact PI: Scott Hazellhurst

HE2AT* | Grant Number: 5US4TW012083-03
 Developing data science solutions to mitigate the health impacts of climate change in Africa: the HE2AT Center
Data Types: Demographic, Surveillance, Clinical/Phenotype, Environmental, Climate, Geospatial
Platform: UCT HPC
Capacity Building: Data Management, Meta-data Principles, Climate/Health Data Analysis, Data Visualisation, Data Sharing
Focus Areas: Climate Change
WITS Health Consortium (PTY), LTD - South Africa

Contact PI: Matthew Chersich

Each Hub:

- Applies data science to develop new solutions for critical health problems
- Brings together multi-disciplinary expertise
- Engages partners from government and private sector



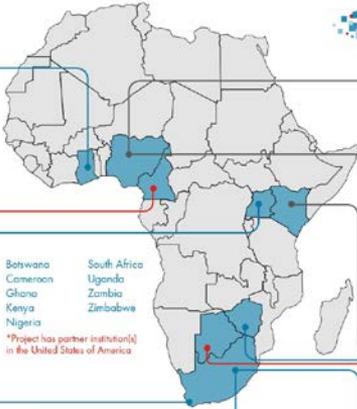
Research Projects

Niche Research Questions

- Congenital heart disease
- Malaria diagnosis & surveillance
- PGRS
- TB Transmission
- Pediatric HIV
- Complex Variant Analysis
- Developmental Disorders
- Early Risk Prediction
- Colorectal Cancer
- Blood cell traits/disease prediction
- Air pollution/Maternal & Neonatal health

RESEARCH PROJECTS



Botswana
Cameroon
Ghana
Kenya
Nigeria

South Africa
Uganda
Zambia
Zimbabwe

*Project has partner institution(s) in the United States of America

DICE*

Grant Number: 1U01ES036147-01

Leveraging Data Science Applications to Improve Children's Environmental Health in Sub-Saharan Africa

Focus Areas: Children's Health, PM2.5 Exposure

Capacity Building: Health Data Science Capacity

University of Cape Coast - Ghana

Contact PI: Adaladza Amegah

CHD AI*

Grant Number: 1U01HL172179-01

Artificial Intelligence assisted echocardiography to facilitate optimal image extraction for congenital heart defects diagnosis in Sub-Saharan Africa

Focus Areas: Congenital Heart Defects, Echocardiography

Capacity Building: Echocardiography Seans

Health Research Foundation - Cameroon

Contact PI: Aminkeng Leke

DS-I Malaria*

Grant Number: 1U01EB035483-01

Automated Mobile Microscopy for Malaria Diagnosis and surveillance in Uganda

Focus Areas: Malaria, Microscopy

Makere University College of Health Sciences - Uganda

Contact PI: Rose Nakasi

BCX-Africa*

Grant Number: 1U01HL172180-01

BCX-Africa: Utilizing data science to evaluate the applicability of blood cell traits polygenic risk scores for disease prediction in Africa

Focus Areas: Blood Cell Traits, Disease Prediction

Capacity Building: Scientific Training

Uganda Virus Research Institute - Uganda

Contact PI: Segun Fatumo

DS-IAFRICA-TB*

Grant Number: 1U01TW012534-01

Tuberculosis in households with infectious cases in Kampala city: Harnessing health data science for new insights on TB transmission and treatment response

Focus Areas: Tuberculosis

Capacity Building: Machine Learning Algorithms

Makere University College of Health Sciences - Uganda

Contact PI: David Kateete

DDD-Africa*

Grant Number: 1U01HD114537-01

Advancing discovery for developmental disorders - expanded analysis of the DDD-Africa resource

Focus Areas: Development Disorders

Capacity Building: Genomic Efforts for Global Health

WITS Health Consortium (PTY), LTD - South Africa

Contact PI: Zané Lombard

INSITE*

Grant Number: 1U01TW012626-01

Innovative data science to impact the TB Epidemic

Focus Areas: Tuberculosis

Capacity Building: Data Infrastructure, Data Harmonisation Platform

University of Cape Town - South Africa

Contact PI: Emma Kalk

ILERA*

Grant Number: 1U01HL172182-01

Integrated models for Early Risk prediction in Africa (ILERA) study

Focus Areas: Cardio Metabolic Disease, Risk Prediction Models

Capacity Building: Public Health Intervention

WITS Health Consortium (PTY), LTD - South Africa

Contact PI: Ananyo Choudhury

GCAP*

Genome-wide characterization of complex variants and their phenotypic effects in African populations

Grant Number: 1U01HG013442-01

Focus Areas: Genetic Variants

Covenant University - Nigeria

Contact PI: Jellie Oyelade

GDRS - Ghana*

Grant Number: 7U01TW012531-02

Geo-enabled detect and respond system for antimalarial resistance in Ghana: GDRS - Ghana

Focus Areas: Molecular Surveillance, Malaria

Institute of Human Virology - Nigeria

Contact PI: Alash'le Abimiku

CHEAQI-MNCH*

Grant Number: 1U01ES036146-01

Characterizing Effects of Air Quality in Maternal, Newborn and Child Health: The CHEAQI-MNCH Research Project

Focus Areas: Air Pollution, Maternal & Neonatal Health

Centre/Sexual HLTH/HIV Aid Res/Zimbabwe

Zimbabwe

Contact PI: Taterida Makanga

DSpace*

Grant Number: 1U01HD114479-01

DSpace: Utilizing Data Science to Predict and Improve Health Outcomes in Pediatric HIV

Focus Areas: HIV, Tuberculosis

Capacity Building: Electronic Health Records, Multi-Omics Data

University of Botswana - Botswana

Contact PI: Goone Retshabisa

Leveraging artificial intelligence/machine learning-based technology to overcome specialized training and technology barriers for the diagnosis and prognostication of colorectal cancer in Africa | Grant Number: 1U01CA287852-01

Focus Areas: Colorectal Cancer

Capacity Building: Diagnostic Tools, Health Infrastructure, Open Source Software

Aga Khan University - Kenya

Contact PI: Mansoor Saleh

ELSI

ETHICAL, LEGAL & SOCIAL IMPLICATIONS PROJECTS



*Project has partner institution(s) in the United States of America

Each ELSI project:

- Conducts bioethical, legal, and social science analyses of new or emerging data science and innovation issues in Africa
- Engage **key stakeholders** and community members
- Collaborates with Research Hubs to **explore emerging areas** of ethical, legal, or social research

BridgELSI* | Grant Number: SU01MH127693-03
Bridging Gaps in the ELSI of Data Science Health Research in Nigeria
Aims: BridgELSI will evaluate current legal instruments, guidelines and frameworks, their implementation, and use these to develop new and innovative governance frameworks to support data science health research in Nigeria.
Center for Bioethics and Research
Nigeria
Contact PI: Temidayo OgunDIRAN

REDSSA*
Grant Number: SU01MH127704-03
Research for Ethical Data Science in Southern Africa
Aims: REDSSA has the overall aims of producing new knowledge in regard to the ethical, legal and social implications (ELSI) of conducting data science research, to develop evidencebased, context specific guidance for the conduct and governance of data science initiatives such as DS-I Africa, and to strengthen the culture of responsible data science in Southern Africa
Stellenbosch University
South Africa
Contact PI: Keymanthri Moodley

DS-I Africa - Law*
Grant Number: SU01MH127690-03
Aims: DS-I Africa Law has a broad jurisdictional scope, involving 12 African nations. Five themes are investigated:
1. modes of informed consent
2. the nature and content of individual and community rights in genomic data
3. the use of persons' geospatial data for public health surveillance
4. cross-border sharing of data
5. the use of data as basis for AI.
University of KwaZulu-Natal
South Africa
Contact PI: Danirch Thaldar

PUBGEM-Africa*
Grant Number: SU01MH127692-03
Public Understanding of Big data in Genomics Medicine in Africa
Aims:
1) to investigate models of public engagement and preparedness for big data use in health; 2) to explore the roles and responsibilities of different stakeholder groups (data providers, data producers, data users, funders, and research ethics committees) regarding intellectual property, patents, and commercial use of genomics big data in health; 3) to investigate public perceptions of big data in health and attitude towards governance of data; 4) to develop data governance frameworks for big data-driven innovation and health in Africa.
University of Cape Town
South Africa
Contact PI: Ambrose Workam

Research Training Hubs



- Each Training Program:
- Trains African faculty
 - Trains next gen African data scientists
 - Develops and implements new MS/PhD curriculum

Research Education Projects

RESEARCH EDUCATION PROJECTS



Ghana
Kenya
Mali

Nigeria
Rwanda

*Project has partner institution(s) in the United States of America

Research Education Projects address very specific training & education challenges identified within the Consortium. E.g, synthetic data generation

DATICAN* | Grant Number: 1UE5E0035490-01
Data Science and Medical Image Analysis Training for Improved Health Care Delivery in Nigeria
Capacity Building: Hackathons, Image Processing, Current Research in Data Science
Focus Areas: Medical Imaging, Data Science
Lagos State University Teaching Hospital - Nigeria
Contact PI: Benjamin Arubisola

GRASP* | Grant Number: 1UESH172183-01
Growing Data science Research in Africa to Stimulate Progress
Capacity Building: Data Science
Focus Areas: Brain Health
College of Medicine, University of Ibadan
Nigeria
Contact PI: Mayowa Owelabi

West Africa Center of Excellence for Data Science Research Education* | Grant Number: 1UESTW012526-01
Capacity Building: Processing Large & Complex Datasets, Dataflow Training
Focus Areas: Data Science Training
University of Sciences, Tech & Tech of Bamako - Mali
Contact PI: Seydou Doumbia

SYNAPSE* | Grant Number: 1UESH172181-01
SYNthetic Healthcare Data Platform for Data Science Training
Capacity Building: Synthetic Healthcare Data, MDClone
Focus Areas: Synthetic Data
National University of Rwanda - Rwanda
Contact PI: David Tumusime

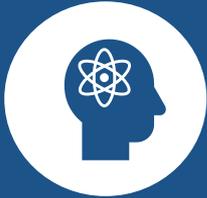
ENEZA* | Grant Number: 1UESTW012539-01
Eneza Data Science: Enhancing Data Science Capability and Tools for Health in East Africa
Capacity Building: Open Science Skills & Tools
Focus Areas: Clinical Practice, Health Research & Policy
International Centre of Insect Physiology
Kenya
Contact PI: Daniel Masiga

DS-ZOOFOOD* | Grant Number: 1UESTW012566-01
Application of Data Science to Build Research Capacity in Zoonoses and Food-Borne Infections in West Africa
Capacity Building: Biostatistics, Phylodynamics, Disease Modelling
Focus Areas: Zoonoses, Food-Borne Infections
College of Health Sciences, University of Ghana - Ghana
Contact PI: Eric Sampene-Dankor



DS-I Africa Accomplishments

Innovation



- 185+ publications on topics like emerging ethical issues in AI and big data in Africa
- developed and deployed a scalable data-sharing gateway (<https://elwazi.org>)
- created an AI Chat Bot that helps researchers navigate data sharing regulations

Education



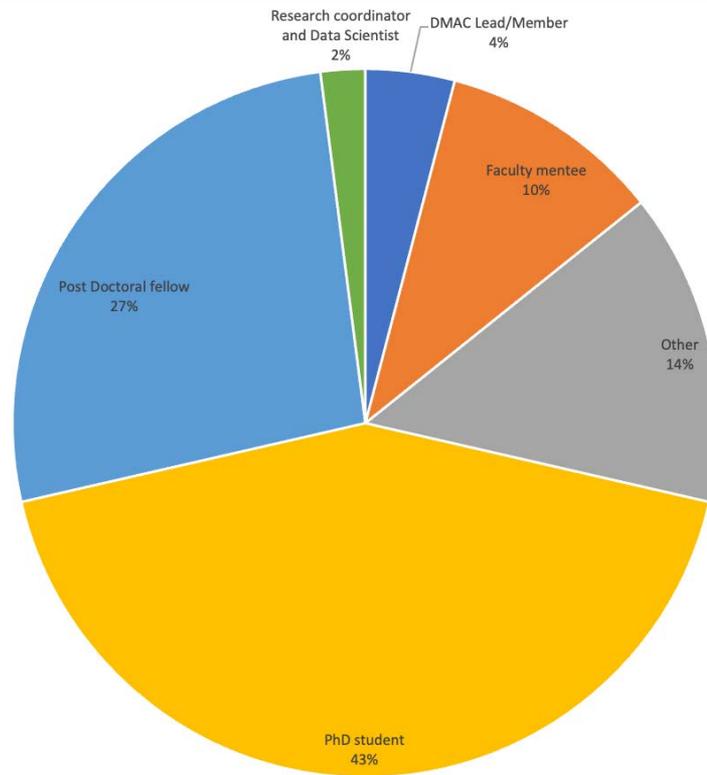
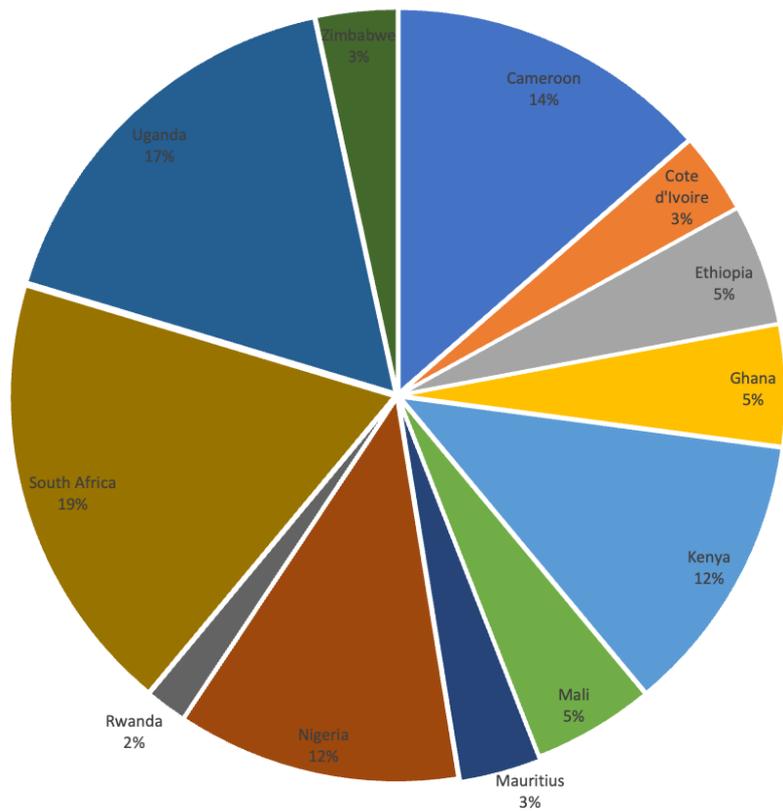
- launched health data science advanced degree programs at 8 African institutions
- training 120 students, 46 faculty, and 11 post-docs, including Ministry of Health staff

Collaboration



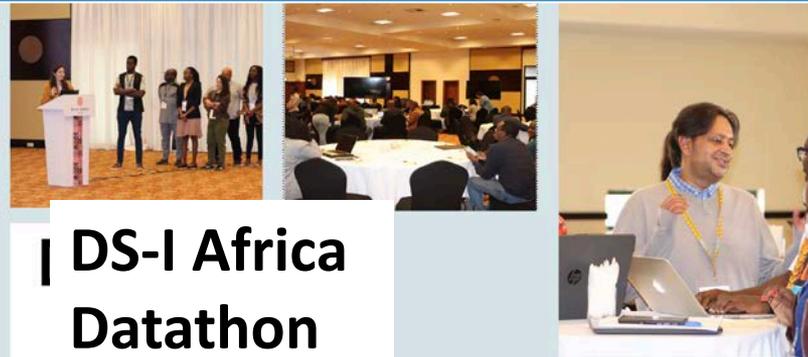
- 200+ partnerships developed and maintained
- engaged with hundreds of data scientists, companies, NGOs, and government staff through networking exchanges, workshops, and presentations

Training

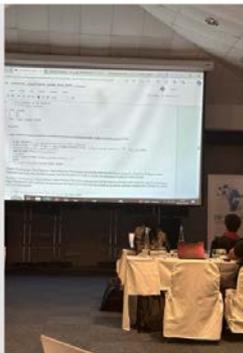


Trainee role and nationality

Training



DS-I Africa Datathon in Kigali, Rwanda



Good day, **Ashaka** folks!
DS-I Africa Data Bootcamp Team announcements, info and agenda
We're excited to let you know that your participation in the **DS-I Africa Data Bootcamp** is confirmed.
We hope you're ready for the event happening over from 1st-27 Nov at the **Kigali Akiba Hotel, Kigali**.
We have been assigned to **GROUP 10**.
Ready for further information regarding the event.
Organizers: DS-I Africa Co-ordinating Centre (aka) DSIF (DSIFA), LLC, MASHITA, LLC, MASHITA Trust, MTT Ltd members
Theme: Exploring Cancer Data
Learning Objectives:

- Maximize the power of breast cancer data to create accurate and insightful predictions
- Participate in hands-on exercises to demystify the black box nature of machine learning models, ensuring your predictions are also understandable
- Learn how to effectively communicate your findings through a guided tutorial on building powerful, data-driven apps

Data sources, tools and infrastructure

- The **MOH45** dataset
- The **MOH45** dataset
- The **MOH45** dataset
- The **MOH45** dataset



DS-I Africa Data Bootcamp Mauritius

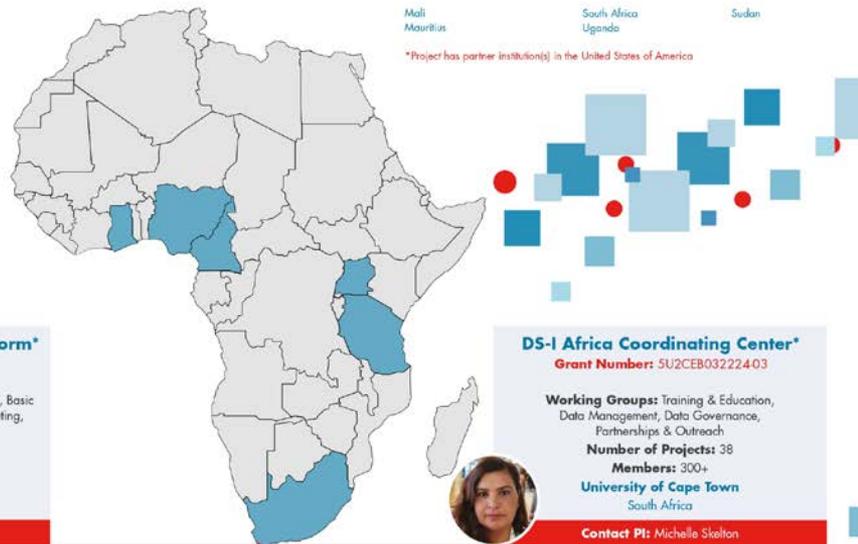
eLwazi ODSP-CC

ELWAZI - ODSP - CC



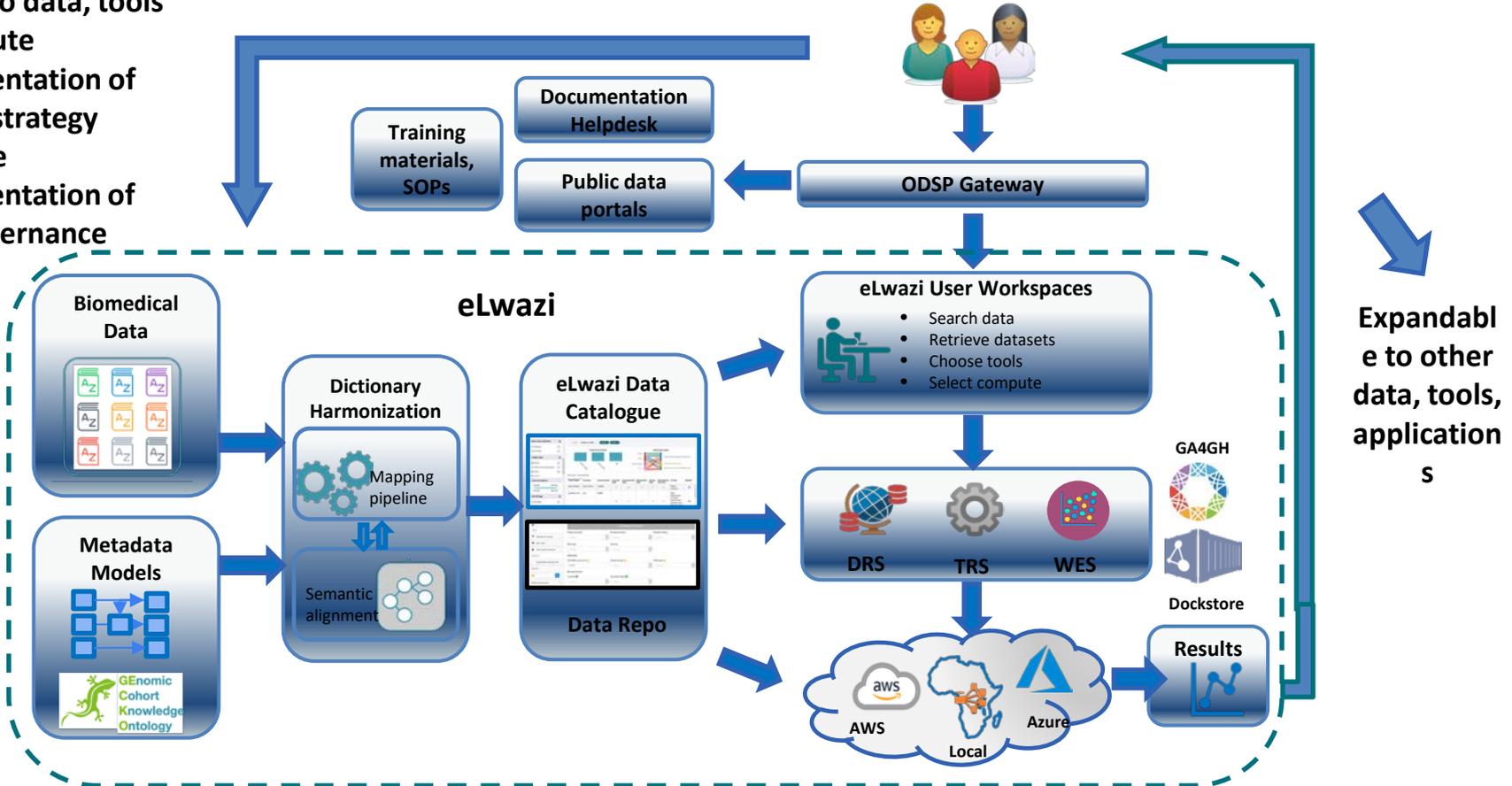
Facilitate collaboration through **community building and platform development:**

- Facilitate working groups and trans-network projects
- Coordinate communication
- Scope potential **partners**
- Enable **discovery** of DS-I Africa data sets
- Access tools and workflows for cross-consortium projects



Overview of the eLwazi ODSP

- Provide support related to data, tools & compute
- Implementation of DMWG strategy
- Facilitate implementation of data governance policies

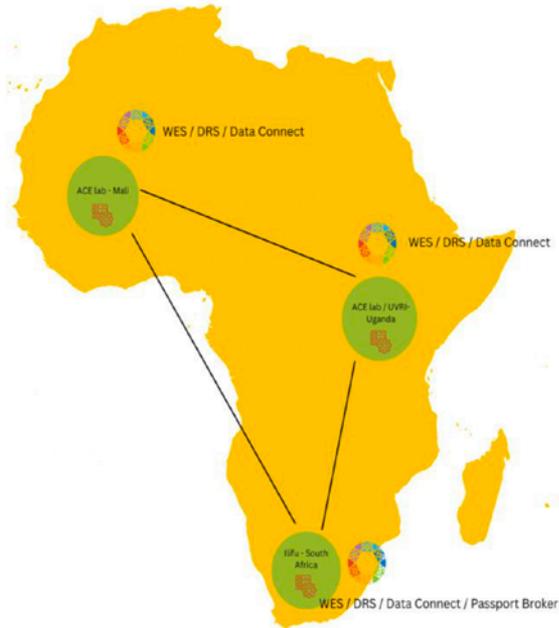


eLwazi Catalogue

eLwazi DS-I Africa									
ENGLISH									
COLUMNS EXPORT									
Project	Name	Category	Type	Countries	Sample Size	Data Use Conditions			
<input type="checkbox"/>	HEZAT	RP1	Air Quality Climate Demographic & Health	Secondary (Derived) dataset	Benin Botswana Burkina Faso Cameroon Côte d'Ivoire Ethiopia Gambia (the) Ghana Kenya Malawi Nigeria Somalia South Africa Tanzania the United Republic of Uganda Zambia Zimbabwe	30,000	Health or Medical or Biomedical Research (HMB)		
<input type="checkbox"/>	HEZAT	RP2	Demographic & Health	Secondary (Derived) dataset	Côte d'Ivoire South Africa	10,000	Health or Medical or Biomedical Research (HMB)		
<input type="checkbox"/>	MADIVA	Verbal Autopsy Dataset	Demographic & Health	Primary dataset	Kenya South Africa		Health or Medical or Biomedical Research (HMB)		
<input type="checkbox"/>	MADIVA	NCD	Demographic & Health	Secondary (Derived) dataset	Kenya South Africa	51,023			
<input type="checkbox"/>	MADIVA	Individual SES indicators	Demographic & Health	Primary dataset	Kenya South Africa				
<input type="checkbox"/>	MADIVA	Household assets	Demographic & Health	Primary dataset	Kenya South Africa				
<input type="checkbox"/>	MADIVA	Individuals	Demographic & Health	Primary dataset	Kenya South Africa				
<input type="checkbox"/>	MADIVA	IndividualEvents	Demographic & Health	Secondary (Derived) dataset	Kenya South Africa				
<input type="checkbox"/>	MADIVA	HouseHolds	Demographic & Health	Primary dataset	Kenya South Africa				
<input type="checkbox"/>	MADIVA	HHMemberships	Demographic & Health	Primary dataset	Kenya South Africa				
<input type="checkbox"/>	MADIVA	Location	Demographic & Health	Primary dataset	Kenya South Africa				
<input type="checkbox"/>	MUDSReH	Modria - Demographics	Demographic & Health	Primary dataset	Kenya Uganda	5,000	Health or Medical or Biomedical Research (HMB)		
<input type="checkbox"/>	MUDSReH	Modria - Images	Image	Primary dataset	Kenya Uganda	5,000			
<input type="checkbox"/>	UZIMA-DS	Health care workers mental health study	Demographic & Health	Primary dataset	Kenya	527	Health or Medical or Biomedical Research (HMB)		
<input type="checkbox"/>	UZIMA-DS	Student survey	Demographic & Health	Primary dataset	Kenya	400	Health or Medical or Biomedical Research (HMB)		
<input type="checkbox"/>	UZIMA-DS	AKDN delivery tool	Demographic & Health	Secondary (Derived)	Kenya Tanzania the United Republic of	13,000	General Research Use (GRE)		

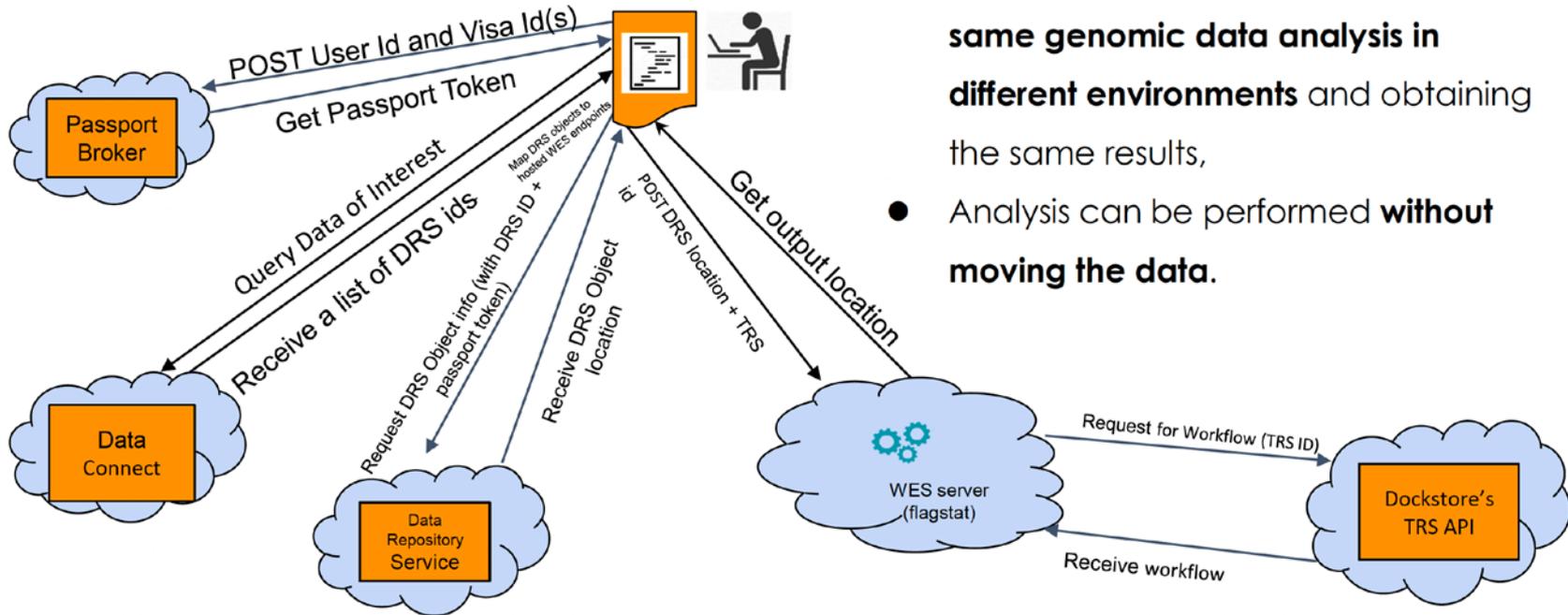
Provides data pointers and metadata on diverse datasets leveraged within the consortium

eLwazi pilot platform for federated analysis



- To show a **proof of concept** of **data sharing and compute** across **African facilities** using **GA4GH standards**
- We used the **1000 Genome data** and selected the **ACE2 region** from the full genome CRAMs
- Data was divided into **four batches** hosted in **DRS servers** at corresponding instances (**Ilifu (SA)**, **ACE lab (Mali)**, **ACE lab (Uganda)** and **UVRI (Uganda)**)
- A user **queries the Data Connect server**, select the **DRS CRAM objects** based on the query and **submits IDs to a WES endpoint**
- The **WES endpoint will process some stats** on the CRAM files and **generate a combined MultiQC report**.

Workflow Execution Service (WES)



- A standard protocol for running the **same genomic data analysis in different environments** and obtaining the same results,
- Analysis can be performed **without moving the data.**

THE 4th
DS-I AFRICA
CONSORTIUM MEETING
16 -22 NOVEMBER 2024

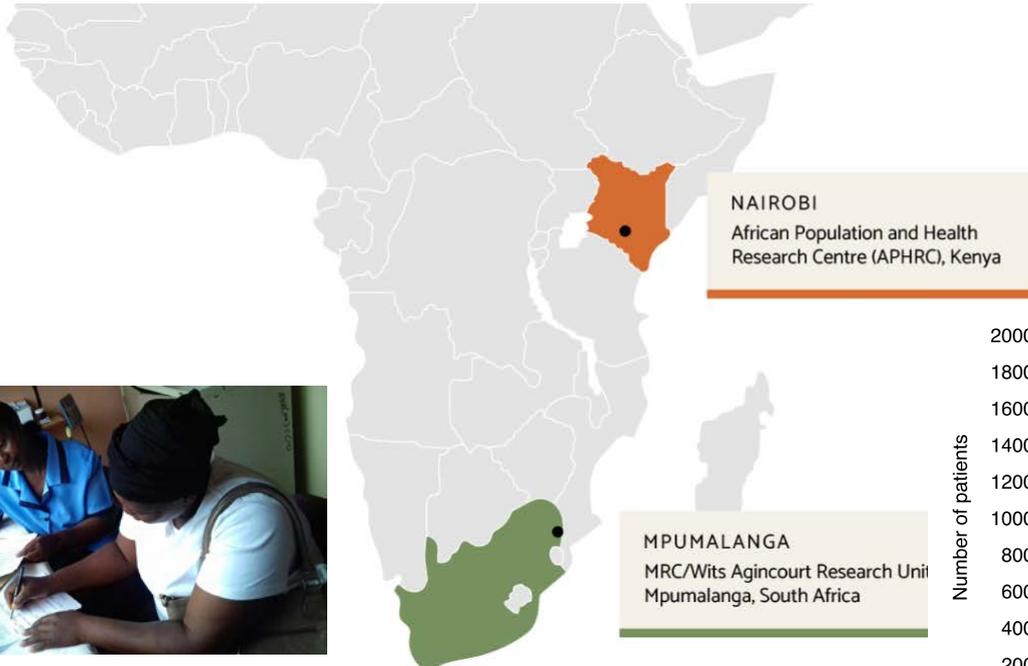
**Stakeholder engagement in
the creation of visualization
tools to support multimorbidity
management at district and
health facility level**

Chodziwadziwa Kabudula
MADIVA Research Hub

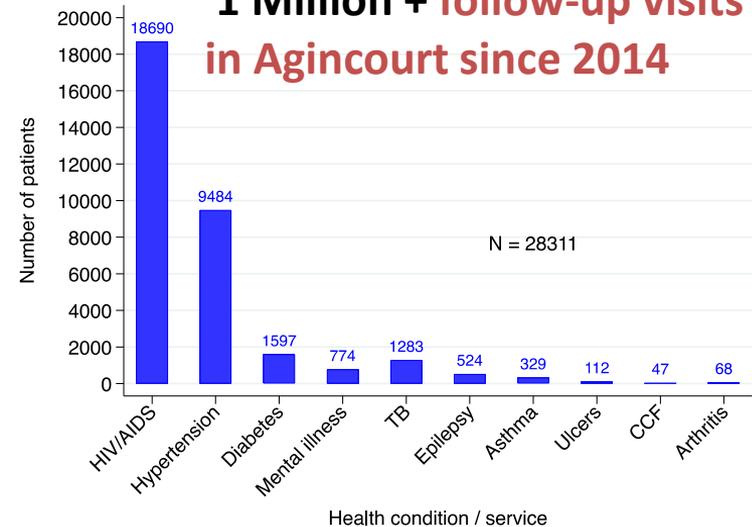


Multimorbidity in Africa: Digital Innovation, Visualization & Application

Research Hub Settings



**1 Million + follow-up visits
in Agincourt since 2014**



Rationale for Stakeholder Engagement

- Ensure dashboard **meet real-world needs**.
- Enhance **usability and adoption** at district and facility levels.
- Promote a sense of **ownership and collective** responsibility.
- Build **stakeholder capacity** for data-driven decision-making.
- Encourage **collaborative feedback loops** for tool improvement.

Engagement Process

1. Engaging with senior health authorities to get political buy-in
2. Health facility visits
3. Requirements gathering workshop with key stakeholders (reference groups)
4. Develop initial wire-frames
5. Review requirements with stakeholders via wireframes
6. Refine wire-frames and design logical flow
7. Design the data layer, functional layer and user interface (UI)
8. Implement design
9. Review initial operational version with stakeholders
10. Deploy, evaluate and refine

Stakeholder Identification

Bushbuckridge, South Africa

- Mpumalanga Provincial Department of Health
- District DOH (Bushbuckridge)
- Health facility administrators
- Healthcare professionals
- Communities

Nairobi, Kenya

- Ministry of Health/National NCD Unit
- Nairobi Metropolitan Services, County Government and City County Health Management Team
- Sub-County Health Management Team
- Health facilities administrators
- Healthcare professional.
- Communities

Requirements gathering workshop with stakeholders

Workshop #1: 30 August 2022 in South Africa

~ 28 participants from Bushbuckridge Health Region

Split into 4 groups: Senior Managers; Nurses; Data Capturers; Operational Managers

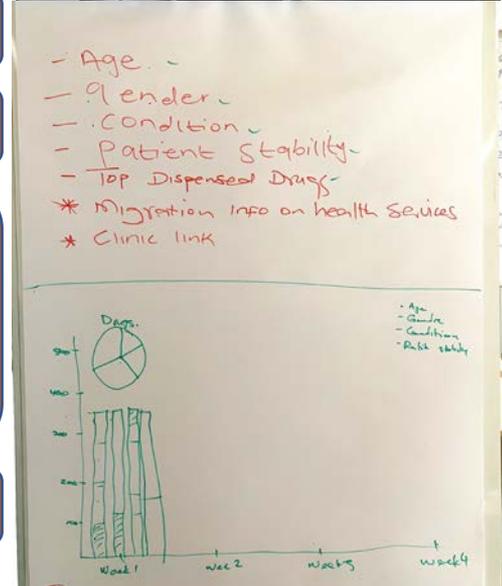
Each group described up to 3 major workflows

With a facilitator, groups brainstormed a list of information used in each workflow

For each element of information:

- What are its units of measurement? Is it quantitative or qualitative? Time-scale? Labels? Max? Min?
- How might it be represented on a dashboard?
- What data would be required to provide this information?
- How important is it? [Essential, useful, nice-to-have]

On a flip chart draw a set of dashboards – present, soon, future



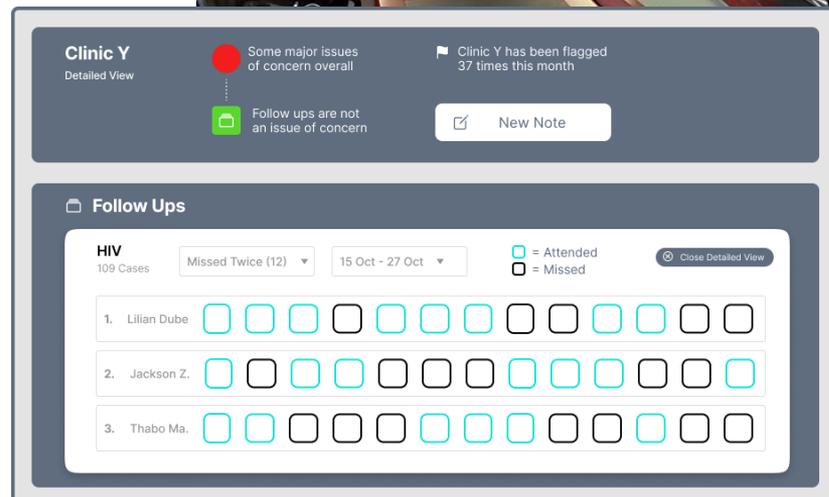
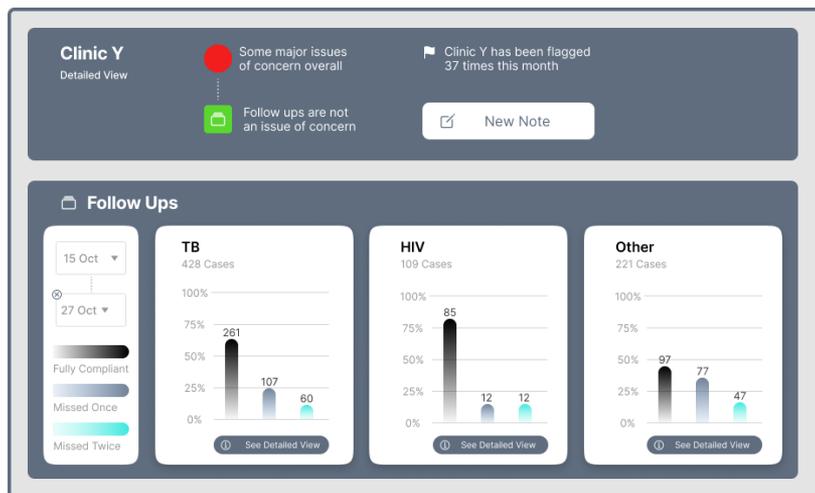
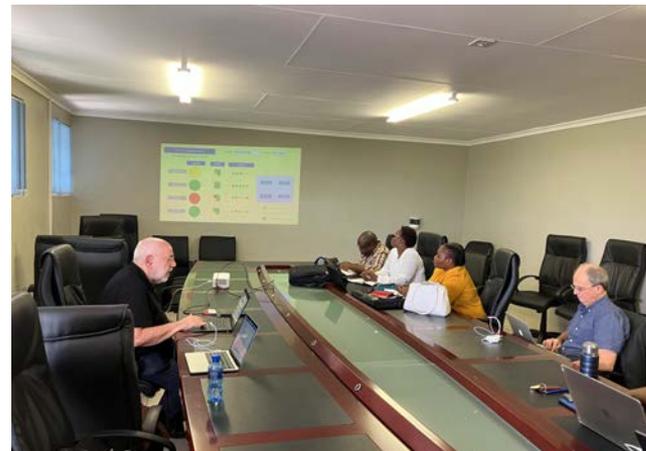
Reviewing Wireframes with stakeholders

Presented initial wireframes to senior management from Health Department – Nov 2022 – Wits Rural

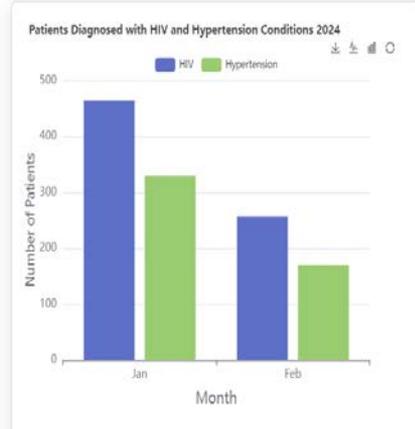
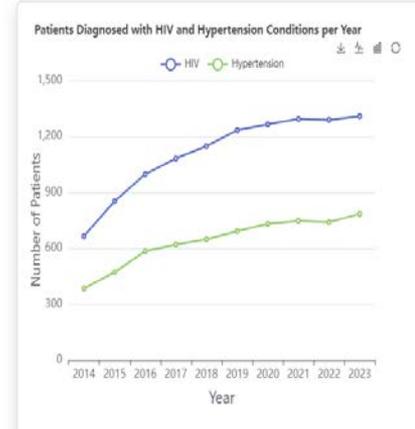
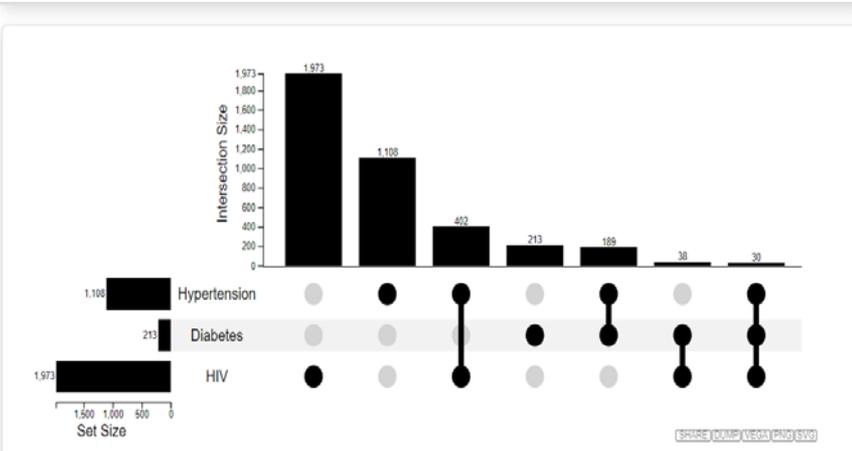
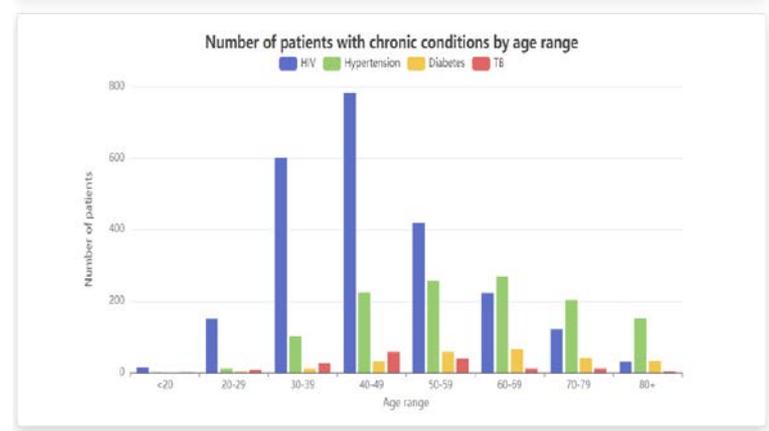
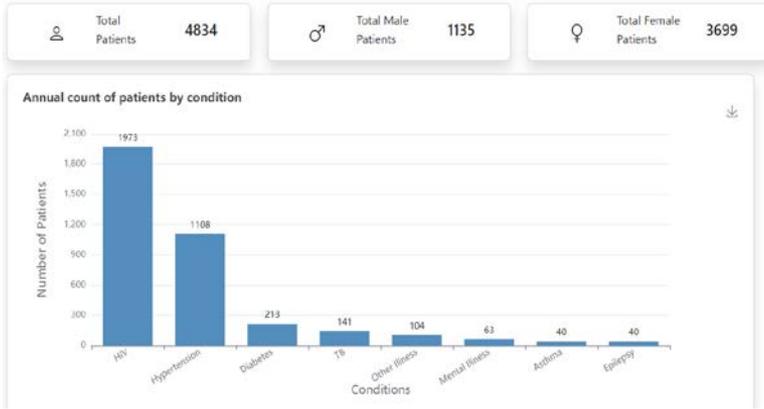
Revised wireframes based on feedback received

Presented revised wireframes to Senior Management Group – February 2023 – Mkhuhlu (near Nelspruit)

Revised wireframes based on feedback received



Reviewing initial operational versions with stakeholders



Challenges and mitigation strategies

Challenges

- Time constraints
- Resource intensive
- Limited technical skills among users

Strategies

- Simplify tools for non-technical users.
- Incentives for participation.
- Ongoing technical support.

DSpace

Differentially Expressed Transcripts in HIV-TB Coinfected Children from East and Southern Africa Reveal Neutrophil Function Enrichment and Small Distinct Transcriptional Signature via Machine Learning

Eric Katagirya¹, Gaone Retshabile^{2*}, Busiswe Mlotshwa², Lesedi Williams², Savannah Mwesigwa¹, Dithan Kiragga³, Mogomotsi Matshaba⁴, Ahmed Alaa^{5*}, Samuel Kyobe^{1*}

¹Makerere University, Kampala, Uganda, ²University of Botswana, Gaborone, Botswana, ³Baylor COE Uganda, ⁴Baylor COE Botswana, ⁵University of California, Berkeley, * Principal Investigators

Introduction

- HIV is the biggest risk factor for TB.

- HIV masks TB presentation causing a diagnostic dilemma.
- The pathophysiology of the two infections is not well understood.
- HIV and *Mycobacteria tuberculosis* infection cause changes in alternative splicing (AS) in macrophages and CD4+ cells (Byun et al., 2020; Kalam et al., 2017).
- Different gene transcripts arise from alternative splicing.
- Differentially expressed transcripts (DETs) may be biomarkers for HIV-TB coinfection.

Study Aim

To identify and characterize differentially expressed transcripts in HIV-TB coinfected children from three Sub-Saharan African countries.

Bell & Noursadeghi, 2018

Methods and Materials

Study sites: Uganda, Botswana, Eswatini

Study population: HIV+ TB co-infected vs HIV mono-infected

Sequence analysis: Bulk RNA sequencing, Standard RNAseq QC Alignment with Salmon Reference - GRCCh37

Downstream analysis: DTE using DESeq2, Pathway and Gene Ontology - Metascape, STRING, Enrichr

Significance threshold: ≥ 2 log₂ fold change, adjusted p-value < 0.05

Results (a)

- Two times more transcripts with significant DTE in Uganda.

Study samples: 97 samples in total
 - Uganda (n=59)
 - Southern Africa (n=38)

Total vs Filtered Transcripts

Volcano plots: Log₂ Fold Change vs -Log₁₀ Adjusted P-Value. Uganda: 185 upregulated, 387 downregulated. Southern Africa: 46 upregulated, 173 downregulated.

Filtered: ≥ 10 reads in ≥ 3 samples

Results (b)

- Upregulated DETs cluster the samples by condition.

Uganda: Volcano plot showing upregulated DETs (red) and downregulated DETs (blue).

Southern Africa: Volcano plot showing upregulated DETs (red) and downregulated DETs (blue).

Pathways Pathways enrichment: Heatmap showing enrichment of biological pathways for upregulated DETs in Uganda and Southern Africa.

- Pathways of DETs are enriched for immune functions – top of which include neutrophil degranulation.
- 10 DETs are shared between Uganda and Southern Africa, including *PRTN3*, *OLFM4*, *LTF*, and *MMP8*.

Results (c)

- Shared DETs are predominantly enriched in secretory compartments.
- Using Machine learning, 11 and 18 features (transcripts) are able to differentiate HIV-TB coinfection from HIV mono-infection in Southern Africa and Uganda respectively.

Southern Africa: Heatmap and ROC curve showing machine learning classification of HIV-TB coinfection.

Uganda: Heatmap and ROC curve showing machine learning classification of HIV-TB coinfection.

Conclusion

- Upregulated DETs in HIV-TB coinfected children are predominantly from genes involved in neutrophil functions among other immune functions.
- Region specific DET based transcription signatures prioritized using machine learning are able to delineate HIV-TB coinfection from HIV mono-infected.

References

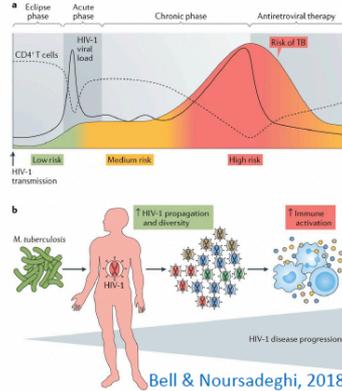
Bell, L. C. K., & Noursadeghi, M. (2018). Pathogenesis of HIV-1 and *Mycobacterium tuberculosis* co-infection. In *Nature Reviews Microbiology* (Vol. 16, Issue 2, pp. 80–90). Nature Publishing Group. <https://doi.org/10.1038/nrmicro.2017.128>

Byun, S., Han, S., Zhang, Y., Planellas, V., & Lee, Y. (2020). The landscape of alternative splicing in HIV-1 infected CD4 T-cells. *BMC Medical Genomics*, 13(5), 1–10. <https://doi.org/10.1186/s12920-020-0680-7>

Kalam, H., Fontana, M. F., & Kumar, D. (2017). Alternate splicing of transcripts shape macrophage response to *Mycobacterium tuberculosis* infection. *PLoS Pathogens*, 13(3), e1006236. <https://doi.org/10.1371/journal.ppat.1006236>

Introduction

- HIV is the biggest risk factor for TB.

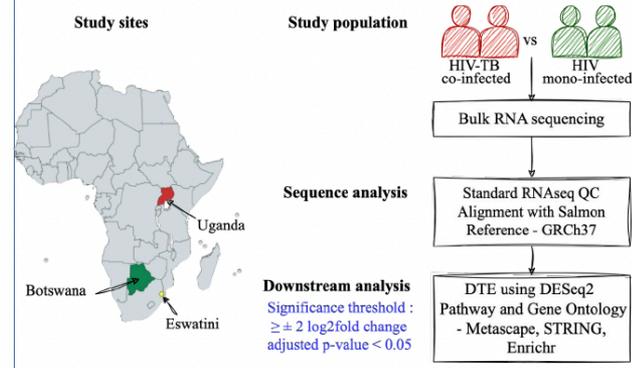


- HIV masks TB presentation causing a diagnostic dilemma.
- The pathophysiology of the two infections is not well understood.
- HIV and *Mycobacteria tuberculosis* infection cause changes in alternative splicing (AS) in macrophages and CD4+ cells (Byun et al., 2020; Kalam et al., 2017).
- Different gene transcripts arise from alternative splicing.
- Differentially expressed transcripts (DETs) may be biomarkers for HIV-TB coinfection.

Study Aim

To identify and characterize differentially expressed transcripts in HIV-TB coinfected children from three Sub-Saharan African countries.

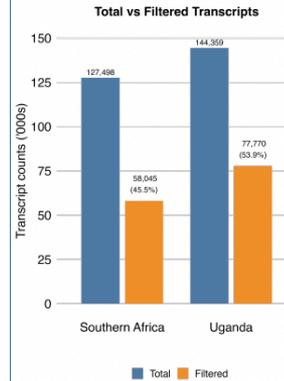
Methods and Materials



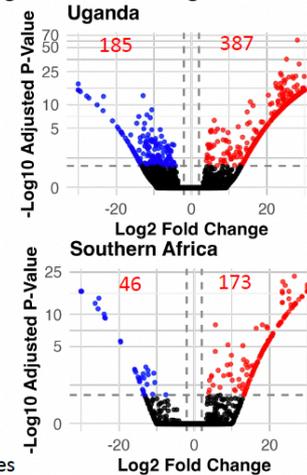
Results (a)

Study samples

- 97 samples in total
- Uganda (n=59)
- Southern Africa (n=38)

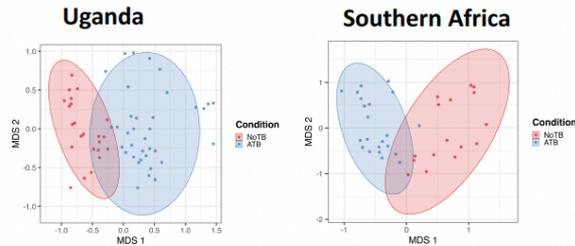


- Two times more transcripts with significant DTE in Uganda.

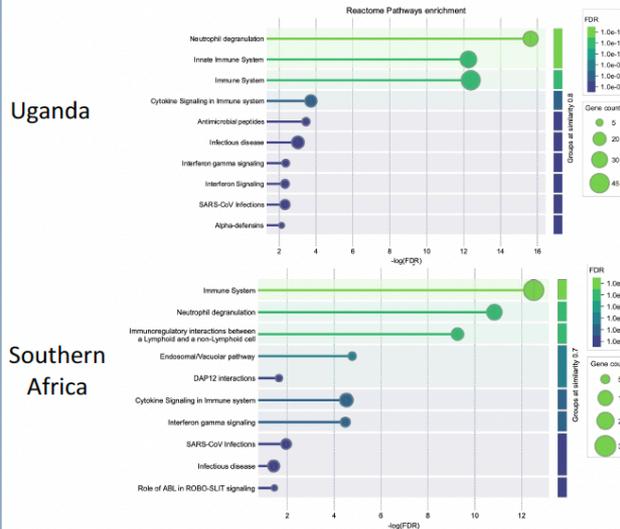


Results (b)

- Upregulated DETs cluster the samples by condition.



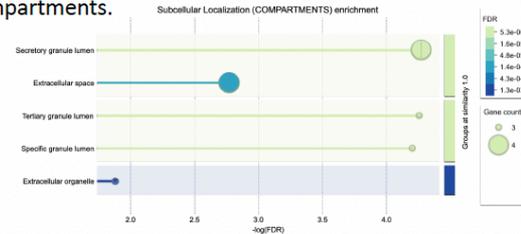
- Pathways of DETs are enriched for immune functions – top of which include neutrophil degranulation.



- 10 DETs are shared between Uganda and Southern Africa, including *PRTN3*, *OLFM4*, *LTF*, and *MMP8*.

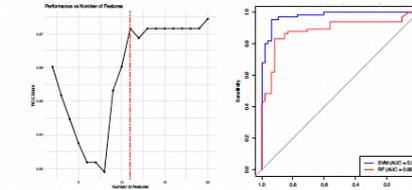
Results (c)

- Shared DETs are predominantly enriched in secretory compartments.

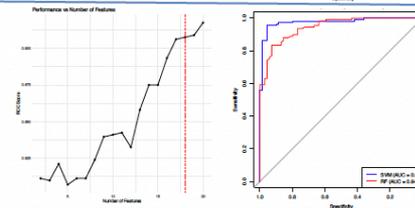


- Using Machine learning, 11 and 18 features (transcripts) are able to differentiate HIV-TB coinfection from HIV mono-infection in Southern Africa and Uganda respectively.

Southern Africa



Uganda

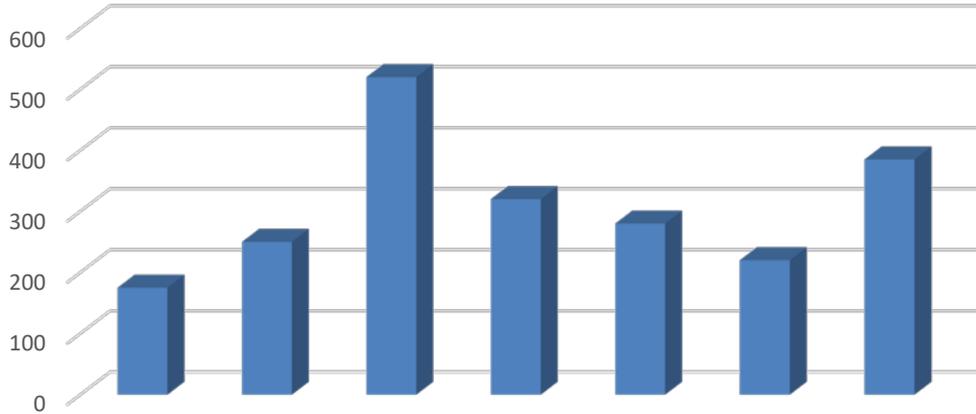


Conclusion

- Upregulated DETs in HIV-TB coinfecting children are predominantly from genes involved in neutrophil functions among other immune functions.
- Region specific DET based transcription signatures prioritized using machine learning are able to delineate HIV-TB coinfection from HIV mono-infected.

Promote partnership and outreach through the Data Science for Health in Africa Networking Exchanges

Networking Exchange Attendance



Networking Exchange (virtual in July 2022) Networking Exchange (in-person in November 2022) Networking Exchange (virtual in May 2023) Networking Exchange (virtual in September 2023) Networking Exchange (in-person in November 2023) Networking Exchange (virtual in April 2024) Networking Exchange (virtual in July 2024)



Consortium Meetings: 4th DS-I Africa Consortium Meeting, Mauritius



Consortium Members



Research Administrators



Data Bootcamp



Steering Committee

<https://dsi-africa.org/dsi-4th-consortium-meeting>

Acknowledgements

University of Cape Town, PI Nicky Mulder
University of the Witwatersrand, PI Scott Hazelhurst
University of the Western Cape/Ilifu, PIs Mattia Vaccari, Rob Simmonds
University of Stellenbosch, PI Tulio de Oliveira
University of Mauritius, PI Shakuntala Baichoo
Uganda Virus Research Institute, PIs Jonathan Kayondo, Daudi Jjingo
University of Khartoum, PI Faisal Fadlemola
USTTB, University of Bamako, PI Seydou Doumbia
Broad Institute, PI Jonathan Lawson, Christine Loreth
EMBL European Bioinformatics Institute, PI Tony Burdett
University of California, Santa Cruz, PI Benedict Paten
University of Chicago, PI Robert Grossman

**Funding: NIH DS-I Africa program, grant
U2CEB032224**



Project Coordinator: Sumir Panji
Working group chairs

eLwazi ODSP PI:
Nicola Mulder

Contact us:
<https://helpdesk.elwazi.org/>



DSpace
Gaone Retshabile
Mogomotsi Matshaba

DS-I Africa Coordinating Centre
Rolanda Julius
Tanian Natus
Tino Muzambi
Francis Agamah
Shonna Smith



Team members

Chodziwadziwa Kabudula, Barry Dwolatzky+, Daniel Maina, Daniel Ohene-Kwofie, Diana Awuor, Faith Kimongo, Isaac Kisiangani, Jacques du Toit, Karabo Maila, Kerry Glover, Patrick Owili, Samuel Iddi, Scott Hazelhurst, Stephen Levitt, F. Xavier Gómez-Olivé.

Funding acknowledgement

National Institutes of Health (Award Number U54 TW 012077)

Fogarty International Center

National Institute of Biomedical Imaging and Bioengineering

Office of Strategic Coordination

Office of Data Science Strategy (ODSS)

South African Medical Research Council.

University of the Witwatersrand.



M A D I V A



THANK YOU



Foster Innovation through the DS-I Africa Ideathon Competition

Team 4: intelliRay: A radiologist in your pocket

intelliRay promises to revolutionize diagnostic practices in underserved regions. The initiative aims to bridge the critical gap in radiological expertise and technology, particularly in remote and resource-limited settings where the burden of respiratory, cardiovascular, and occupational diseases is most acute.

Team 1: Omics Data Harmonization – Developing a big data management pipeline using machine learning

Proposed a machine learning framework for omics data harmonization

Team 2: SAVE ME mobile phone application for road traffic crash response in Cameroon

The "Save Me" app aims to revolutionize emergency response to road traffic injuries by leveraging technology to provide immediate first aid guidance, optimize ambulance routes, and facilitate rapid communication between the public and healthcare providers.

Team 7: Integrative Approach to Predicting Fasting Blood Glucose Levels and Developing a Community-Driven Mobile Application for Diabetes Prevention and Management

Develop an app to predict diabetes



Mamadou D. Coulibaly

University Clinical Research Center, Bamako, Mali
West Africa Center of Excellence for Data Science
Research Education

Bioinformaticians/Data Manager

"This ideathon allowed me to work with a multidisciplinary team to develop a project idea. It was a very enriching exercise. I appreciated the fact that we had enough time to work and we benefited from the assistance of experienced mentors. It is time to thank the entire team of the organization, the mentors and DS-I-Africa for taking this initiative."



Dativa Pereus

University of Nairobi - Kenya and Muhimbili University
of Health and Allied Sciences - Tanzania

Genomics/Bioinformatics

"What I liked about the Ideathon was that the Ideathon process created a space to connect and engage with people from different disciplines and localities, and we committed ourselves to making it possible."



Osolale Olayinka

Department of Biological Science, Elizade University,
Ilara-Mokin, Ondo State, Nigeria

Microbiologist

"What I like about the ideathon: I love that we could do a transborder collaboration without the barrier or limitation of languages, time differences, and busy schedules."



Cherlynn Dumbura

Heat and Health African Transdisciplinary Center
(HEZAT Center)
Centre for Sexual Health and HIV AIDS Research
Zimbabwe (CASHAR)
Place Alert Labs, Midlands State University

Data Science and Biostatistics

"The DS-I Africa Ideathon offered me a valuable opportunity for multidisciplinary team collaboration and access to mentors who provided invaluable guidance in refining our initial idea."



Kaleab Woldemariam

Udacity (student), Addis Ababa University, Ethiopia
(Alumnus)

Science (MSc in Remote Sensing and
GIS, MA in Environmental Studies),
Programming in Kotlin (Android),
Python, SQL, JavaScript and
AI/Machine Learning (Certificates)

"We appreciated the diversity of innovation in Data Science Africa projects. We are happy that we won 3rd runner up! The Judges' and Mentors' suggestions were valuable and positive. Our idea impacts > 300 million diabetes patients worldwide. The Ideathon was a great catalyst for cooperation among Africans."



Theresia Mkenda

Mwenge Catholic University, Tanzania

Mathematical and Statistical Modelling
(PhD)

"The way they first prepared us through a series of trainings before asking us to submit a proposal."



Verrah Otiende

USIU-Africa

Computational Statistician | Data
Science Educator

"The creativity and innovative thinking that shaped each phase of our project were truly standout aspects. The judges' and mentors' insights were invaluable, and it was rewarding to see our project's relevance and impact recognized. This experience underscored the power of collaboration and the effectiveness of a community-driven approach in tackling challenges."



Rukayat Tijani

Lagos State University

AI/ Machine Learning

Next Virtual Networking Exchange – April 30, 2025

■ Feature and promote DS-I Africa Projects

■ Feature and promote Pilot Projects



Data Science for Health in Africa Virtual Networking Exchange

30 April
1 2025

[READ MORE](#)



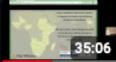
RESOURCES: RECORDINGS & SLIDES



Fourth Virtual Networking Exchange # ×

DS-I Africa - 1 / 11

↺ ↻ ⋮

-  **DS-I Africa Virtual Networking Exchange - DS-I Africa...**
DS-I Africa
1:47:30
-  **DS-I Africa Virtual Networking Exchange Networking Round ...**
DS-I Africa
34:09
-  **DS-I Africa Virtual Networking Exchange Networking Round ...**
DS-I Africa
9:13
-  **DS-I Africa Virtual Networking Exchange Networking Round ...**
DS-I Africa
24:24
-  **DS-I Africa Virtual Networking Exchange Networking Round ...**
DS-I Africa
40:35
-  **DS-I Africa Virtual Networking Exchange Networking Round ...**
DS-I Africa
35:06
-  **DS-I Africa Virtual Networking Exchange Networking Round ...**
DS-I Africa
35:14
-  **DS-I Africa Virtual Networkin...**

Event Agenda Participating Organization **Presentation Slides**

▾ 1) Welcome and Overview of Networking Exchange

[DS-I Africa Coordinating Center/NIH Presentation Slides](#)

▸ 2) Overview of DS-I Africa

▸ 3) Brief Overview of DS-I Africa Partnerships & Outreach Working Group

▸ Networking Round 1 - Room 1: Health Research Centers 1

▸ Networking Round 1 - Room 2: NGO/Non Profit

▸ Networking Round 1 - Room 4: Open Science Network

▸ Networking Round 1 - Room 5: Health Research Centers 2

▸ Networking Round 2 - Room 1: NGO/Non Profit

▸ Networking Round 2 - Room 2: Health Research Centers 3

▸ Networking Round 2 - Room 3: Private Sector

▸ Networking Round 2 - Room 4: Global Networks

▸ Networking Round 2 - Room 5: Health Research Centers 4

▸ Networking Round 3 - Funder's Session

THE 4th
DS-I AFRICA
CONSORTIUM MEETING

16 -22 NOVEMBER 2024

Innovative Data Science to impact the TB Epidemic (INSITE)

Progress with Data Platforms

Presented by: Emma Kalk
Centre for Integrated Data & Epidemiological Research
School of Public Health
University of Cape Town



Setting: Western Cape Province, South Africa

TB annual prevalence: 681/100,000

TB inc. in women living with HIV:

Pregnancy: 1127/100,000 person-years

Post-partum: 1447/100,000 person-years

Antenatal HIV prevalence: 16.3%

Poorly performing TB programme

Drug-resistant TB:

New & repurposed drugs are routine in pregnancy



Ndlovu, DHB 2024; National TB Prevalence Survey 2021; Sugarman, NEJM. 2014; Odayar, IJTBLD. 2018



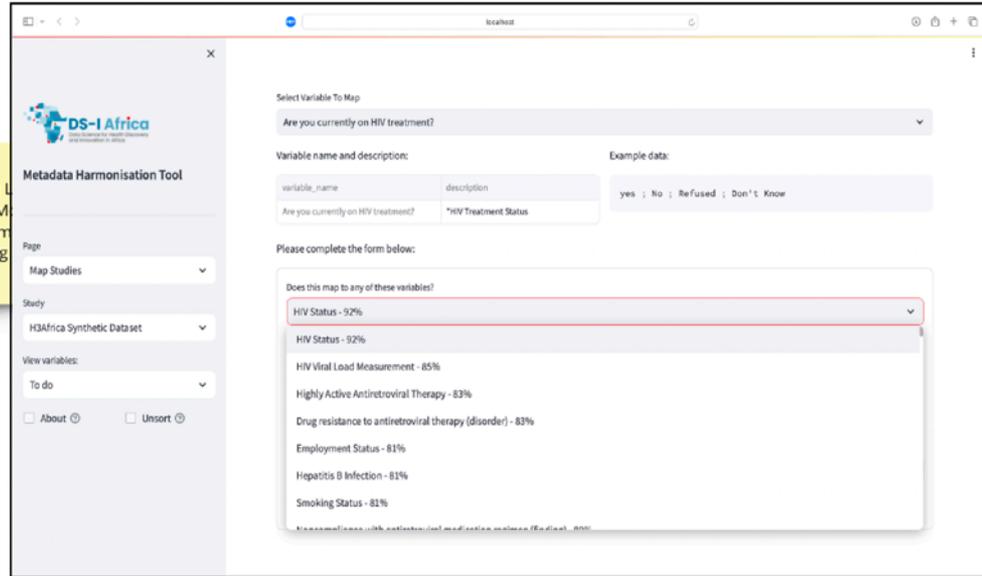
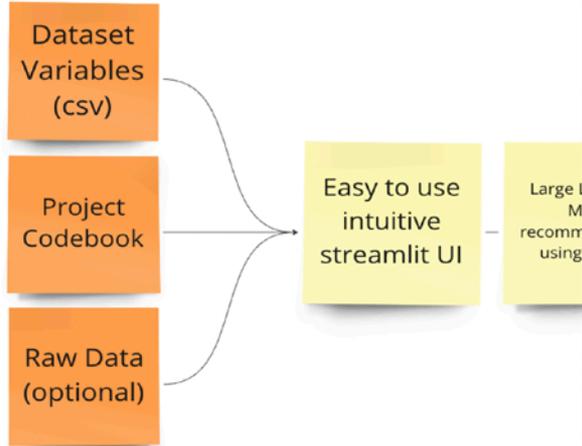
Western Cape Provincial Health Data Centre



Metadata Harmonization Tool



Inputs



Developed by HE2AT project, being further developed with eLwazi to make it more generic & more widely accessible

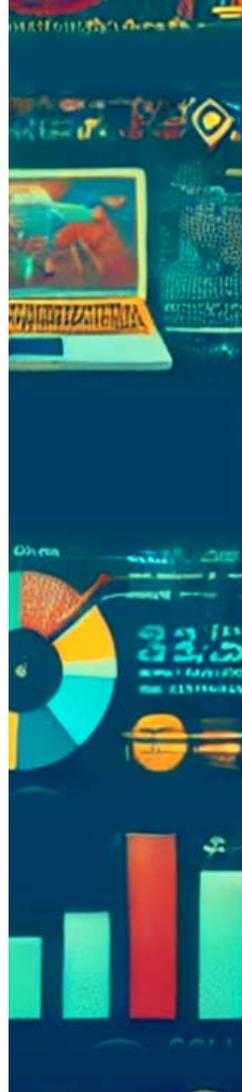
PHDC

Health information exchange

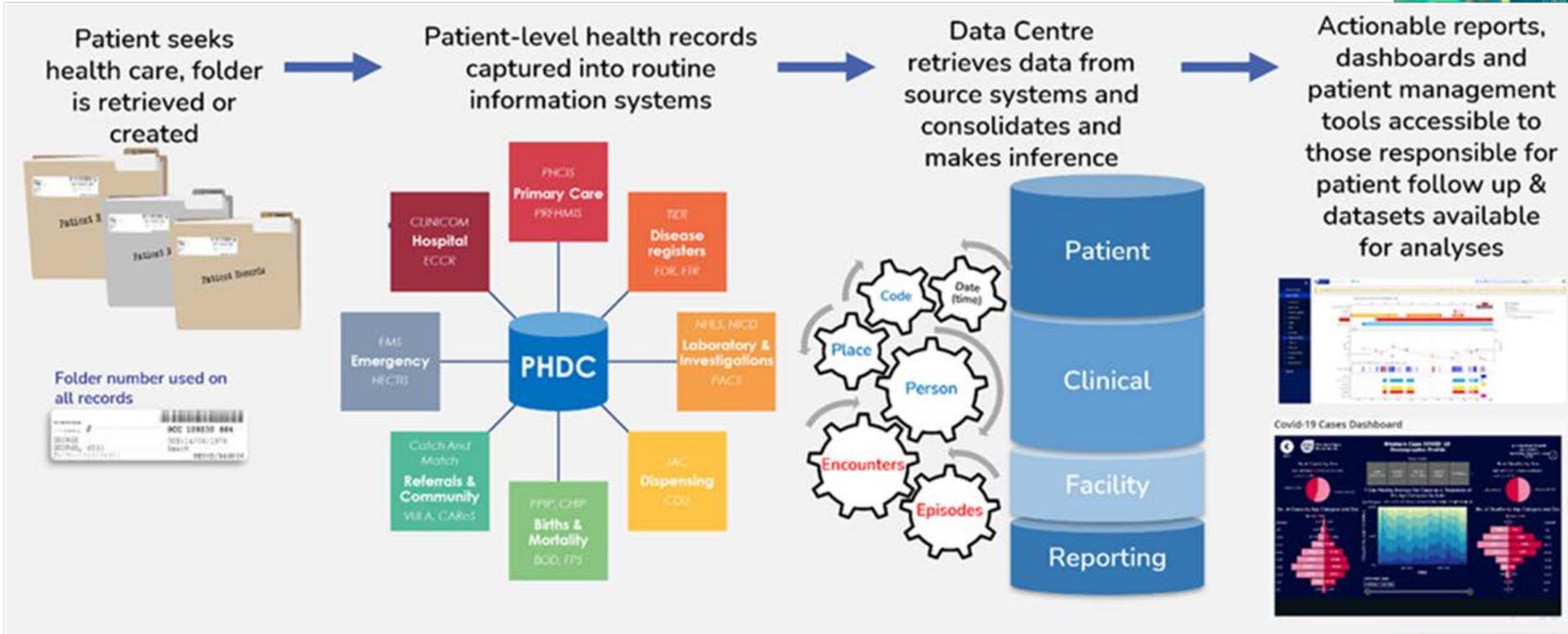
Multi-evidence approach

Manage large datasets to support visualization

Multiple stakeholders

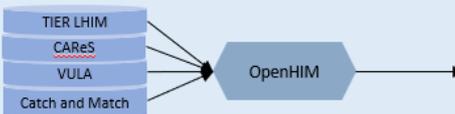


Data Pathways



Sources Imported

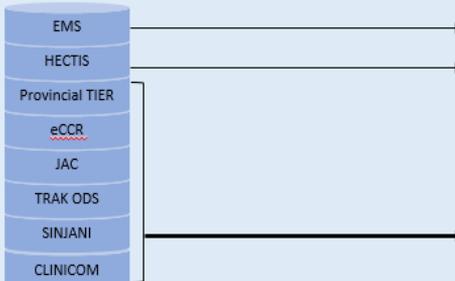
Data sent as FHIR documents via HIE, e.g. [OpenHIM](#)



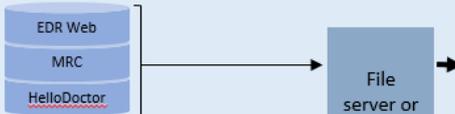
Data pushed into SQL server by source system



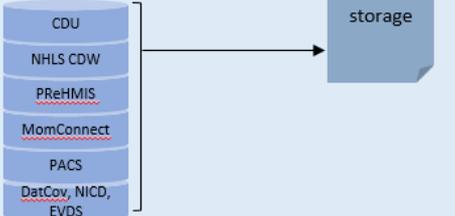
Linked Servers accessed via SQL server or Azure Data Factory



Manual file retrieval and drop onto file server

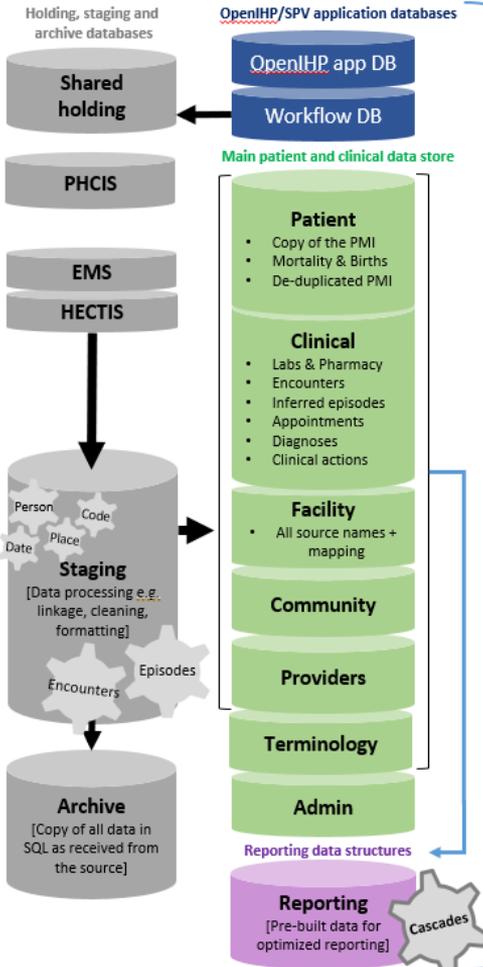


Automatic retrieval from FTP, email, or API

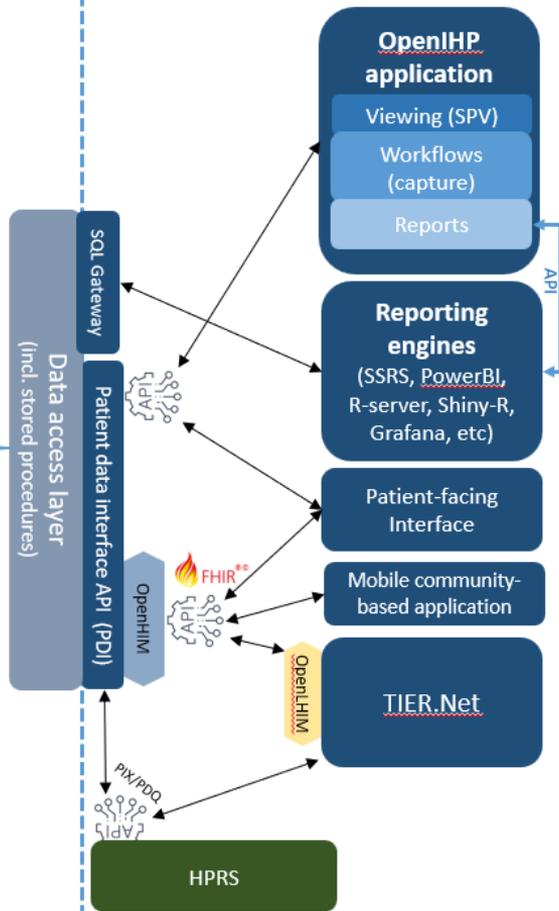


PHDC Databases

(Behind SITA firewall)



Software Applications



Uses





Probabilistic inference



Proportion of projects using AI by project type

