Breakout Session 1: Track B

MUST Data Science Research Hub (MUDSReH) - Democratized Trusted Research Environment (dTRE)

Dr. William Wasswa
Co-PI Admin Supplement, Mbarara University of Science and Technology
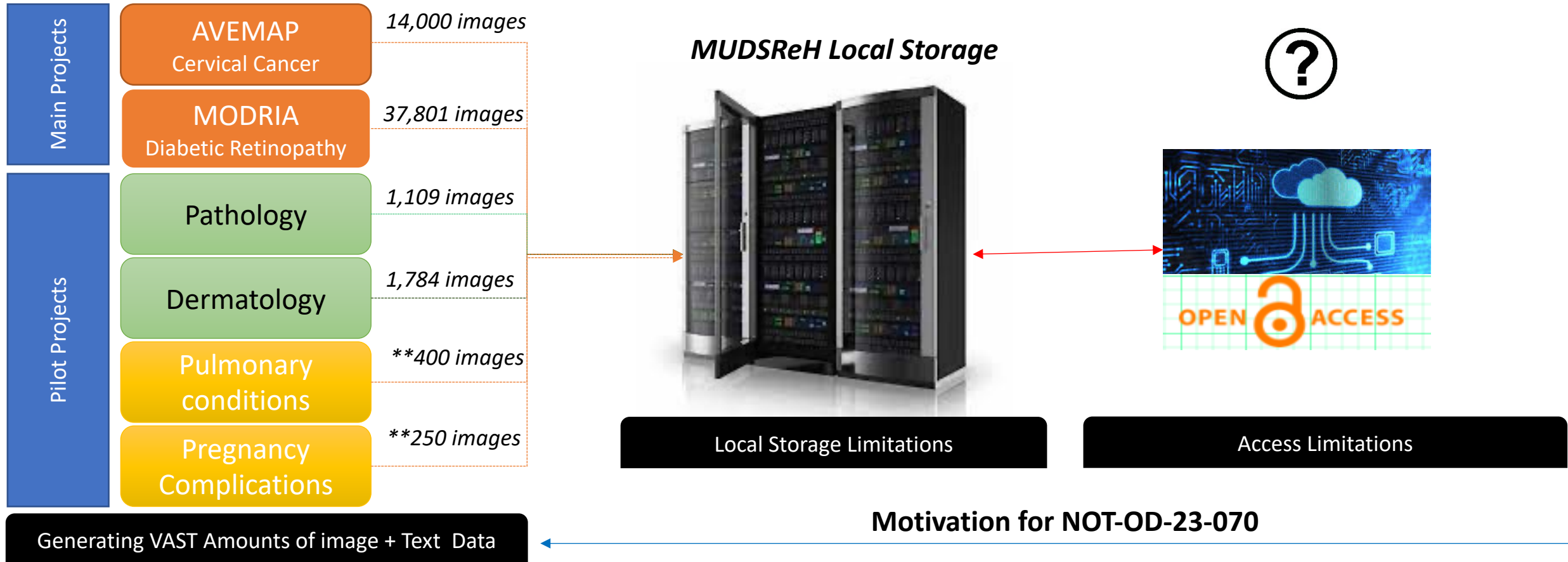
MUST Data Science Research Hub (MUDSReH) – Democratized Trusted Research Environment (dTRE)

17 - 18 Jan 2024:

Democratized Trusted Research Environment (dTRE) - Motivation

Main Projects

| AVEMAP Cervical Cancer | 14,000 images |
| MODRIA Diabetic Retinopathy | 37,801 images |

Pilot Projects

Pathology	1,109 images
Dermatology	1,784 images
Pulmonary conditions	**400 images
Pregnancy Complications	**250 images

MUDSReH Local Storage

Local Storage Limitations

OPEN ACCESS

Access Limitations

Generating VAST Amounts of image + Text Data

Motivation for NOT-OD-23-070

Grant No. U54TW012043

MGH 1811

Democratized Trusted Research Environment (dTRE) - Motivation

MUDSReH
MBARARA UNIVERSITY DATA SCIENCE RESEARCH HUB

- Images + Bio data
- De-identified Data
- Normalised Datasets

MUDSReH Local storage

Promote Research and Data Utilisation
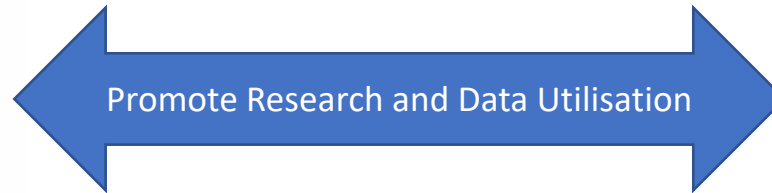
- Images + Bio data
- De-identified Data
- Normalised Datasets

Data availed to Public

OPEN ACCESS

Local Storage Limitations

- Power Instability
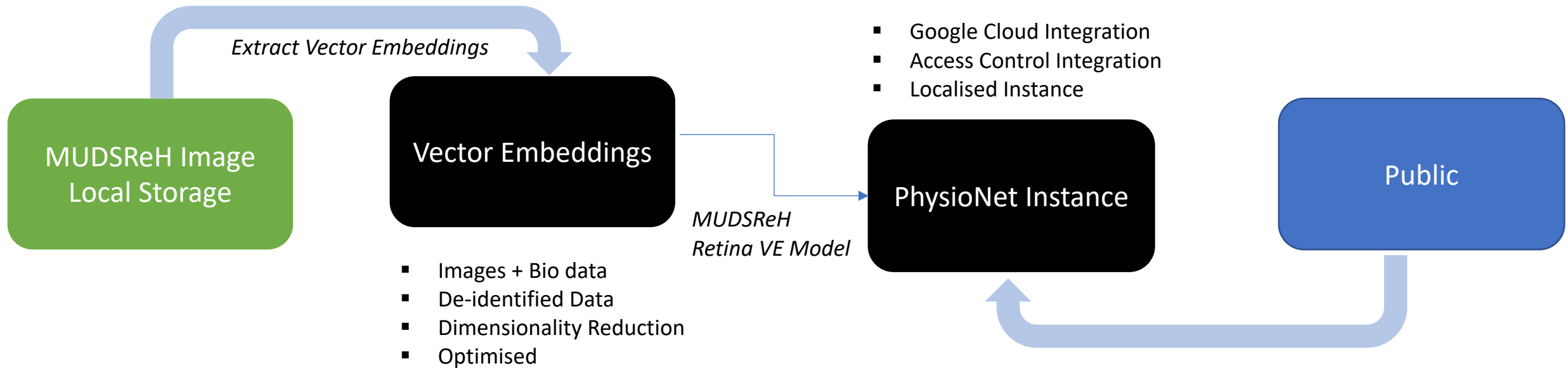- Network Instability
- Internet Access
- Storage capacity
- Compute()

dTRE

Access Limitations

- Save storage space
- Reduced memory req.
- Efficient processing
- Faster data retrieval
- Improved privacy
- Reduced pre-processing
- Reduced bandwidth req.

- Safe Data
- Safe Projects
- Safe People
- Safe Settings
- Safe Outputs
- FAIR Data principles

MGH 1811

Democratized Trusted Research Environment (dTRE) - Design

Extract Vector Embeddings

MUDSReH Image Local Storage

Vector Embeddings

- Images + Bio data
- De-identified Data
- Dimensionality Reduction
- Optimised

MUDSReH Retina VE Model

- Google Cloud Integration
- Access Control Integration
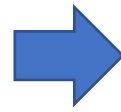- Localised Instance

PhysioNet Instance

Public

Democratized Trusted Research Environment (dTRE) - Design

MUDSReH Retinal Images + Metadata

MUDSReH DMAC + MIT Vector Embeddings Team

PhysioNet
The Research Resource for Complex Physiologic Signals

- ✓ Develop a robust vector embedding model for retina images - Completed
- ✓ Test the Vector embedding Model on Brazilian Dataset - Completed
- ✓ Train the Vector Embedding on MUDSReH Retinal Images – Ongoing – expected to be completed by end of February

- ✓ Create a demonstrator environment using publicly available PhysioNet Data in Google Cloud with novel access control – On Going
  - ✓ Creating MUDSReH project on PhysioNet - Completed
  - ✓ GCP Integration – On Going
  - ✓ Access control integration – On Going
- ✓ Transfer the PhysioNet dTRE infrastructure to the MUDSReH data ecosystem – On Going

Grant No. U54TW012043

Democratized Trusted Research Environment (dTRE) – Expected Outcomes

1) An Optimized Vector Embedding Model for Retinal Images for an African Population

- ✓ Increase Knowledge about Vector Embeddings in Retinal Images– Through a manuscript
- ✓ Increased Utilisation of Retinal Images – Through a datathon

2) A secure African Based PhysioNet Instance

- ✓ Increase Knowledge about Trusted Research Environments – Through a manuscript
- ✓ Increased Utilisation of MUDSReH Images – Through a datathon
- ✓ Increased development of ML Models using MUDSReH Datasets – Through Access to GCP Compute infrastructure

Grant No. U54TW012043

THANK YOU

Email: mudsreh@must.ac.ug | Website: www.must.ac.ug