

DS-I Africa Open Data Science Platform

Nicky Mulder

University of Cape Town

NIH DSI-Africa

- Harnessing Data Science for Health Discovery and Innovation in Africa
 - **Research Hubs:** Advance and demonstrate feasibility of data science research and innovation to improve health in Africa
 - **Training:** Increase capacity for data science research in Africa
 - **ELSI Research:** Explore Ethical, Legal, and Social Implications of data science research from an African perspective and contribute to policy discussion on the continent
 - **Open Data Science Platform & Coordination Center:** Facilitate the development of a trans-African network of data scientists

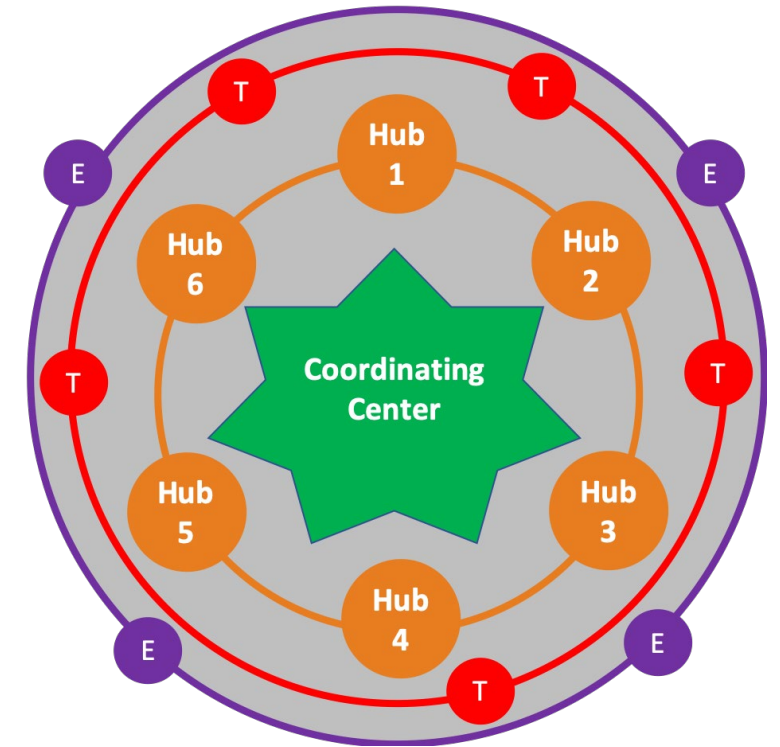


Image and text extracted from: <https://commonfund.nih.gov/africadata>

Funded project titles (1)

Research hubs

- Role of Data Streams In Informing **Infection Dynamics** in Africa- INFORM Africa
- UZIMA-DS: **Utilizing health Information** for Meaningful impact in East Africa through Data Science
- Developing data science solutions to mitigate the **health impacts of climate change** in Africa: the HE2AT Center
- Harnessing Data Science to Promote Equity in **Injury and Surgery** for Africa
- Combatting **AntiMicrobial Resistance** in Africa Using Data Science (CAMRA)
- MADIVA (**Multimorbidity in Africa**: Digital innovation, visualisation and application)
- MUST Data Science Research Hub (MUDSReH)

Funded project titles (2)

ELSI:

- Research for Ethical Data Science in Southern Africa (REDSSA)
- Bridging Gaps in the ELSI of Data Science Health Research in Nigeria (BridgELSI)
- DS-I Africa - LAW
- Public Understanding of Big data in Genomics Medicine in Africa (PUBGEM-Africa)

7 Training grants on different aspects of data science & health

Open Data Science Platform & Coordinating Centre

Open Data Science Platform (ODSP) aim

- To develop an African Open Data Science Gateway, including the eLwazi platform and associated resources, to support the Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) consortium and beyond. eLwazi, will be a flexible, scalable Open Data Science Platform enabling the implementation of data science for health, that is relevant to the African context.

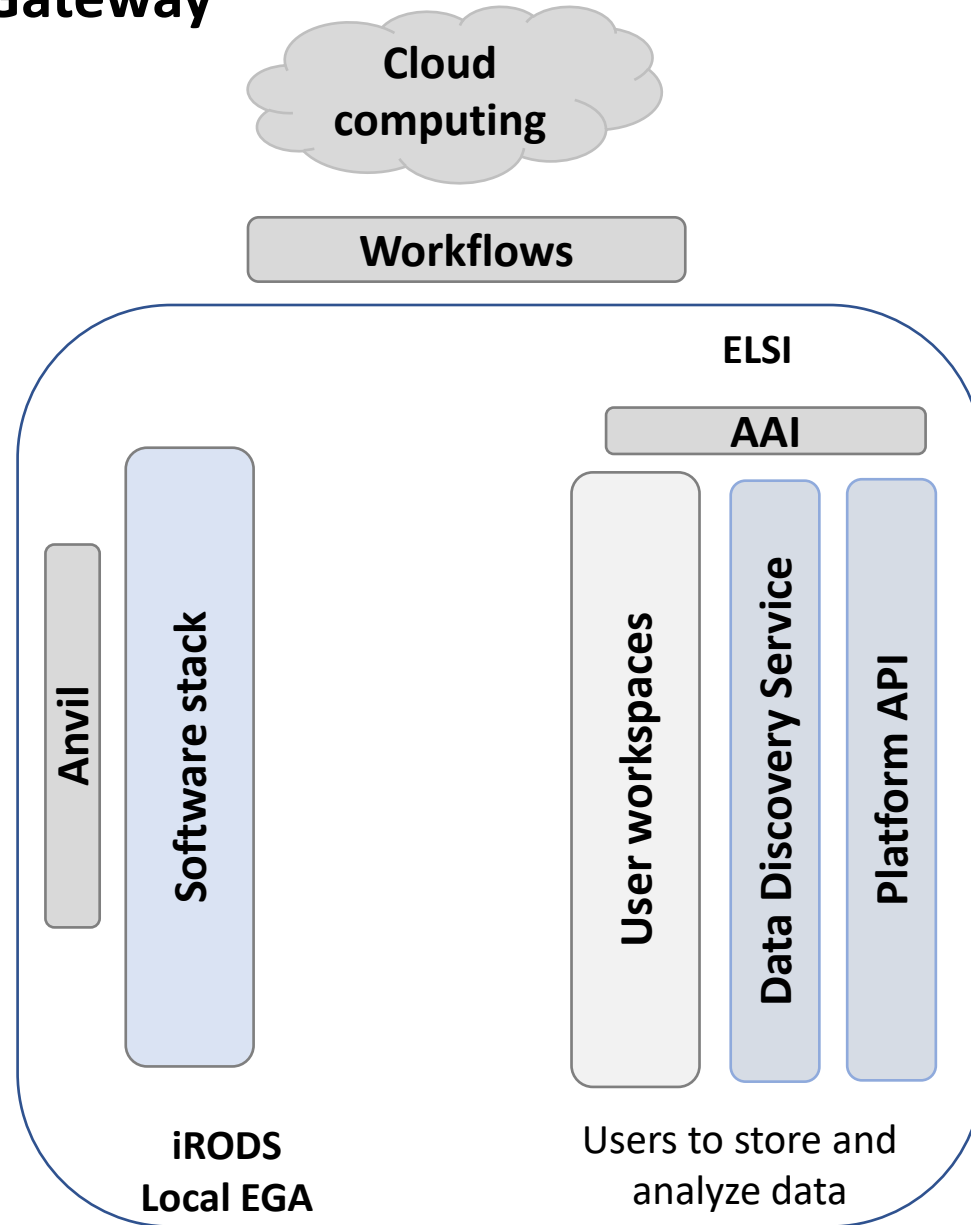
Ulwazi is the Xhosa word meaning “knowledge” or “information”,
and Olwazi means big rock in Luganda



eLwazi

eLwazi Data Science Gateway

AIM 1: Build Infrastructure



Cloud and local computing hosts

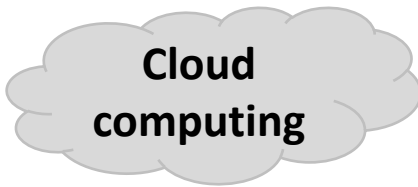
Security and access (AAI)

Aim 2. Access to data and resources

- Collect data types
- Reference datasets
- Metadata stds and curation, FAIR
- Data harmonization
- Cohorts atlas, catalogue
- Linking datasets
- Data Portals
- Other resources: SOPs, protocols, training materials

eLwazi Data Science Gateway

AIM 2: Data, metadata, catalogues



Workflows

Public



Metadata catalogue

Transform data into usable form
DB vs KB

Data portals

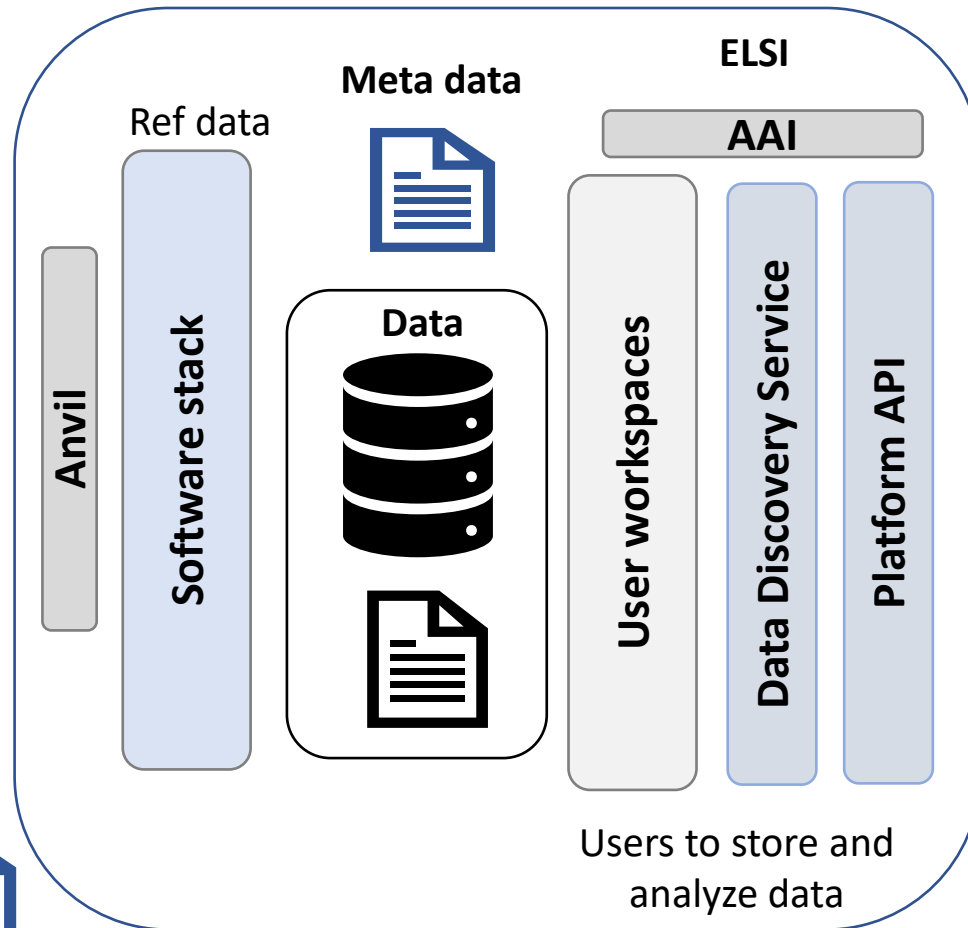
- AGVD
- APDB
- AMP
- Beacon



Training content



Pathogens Protocols

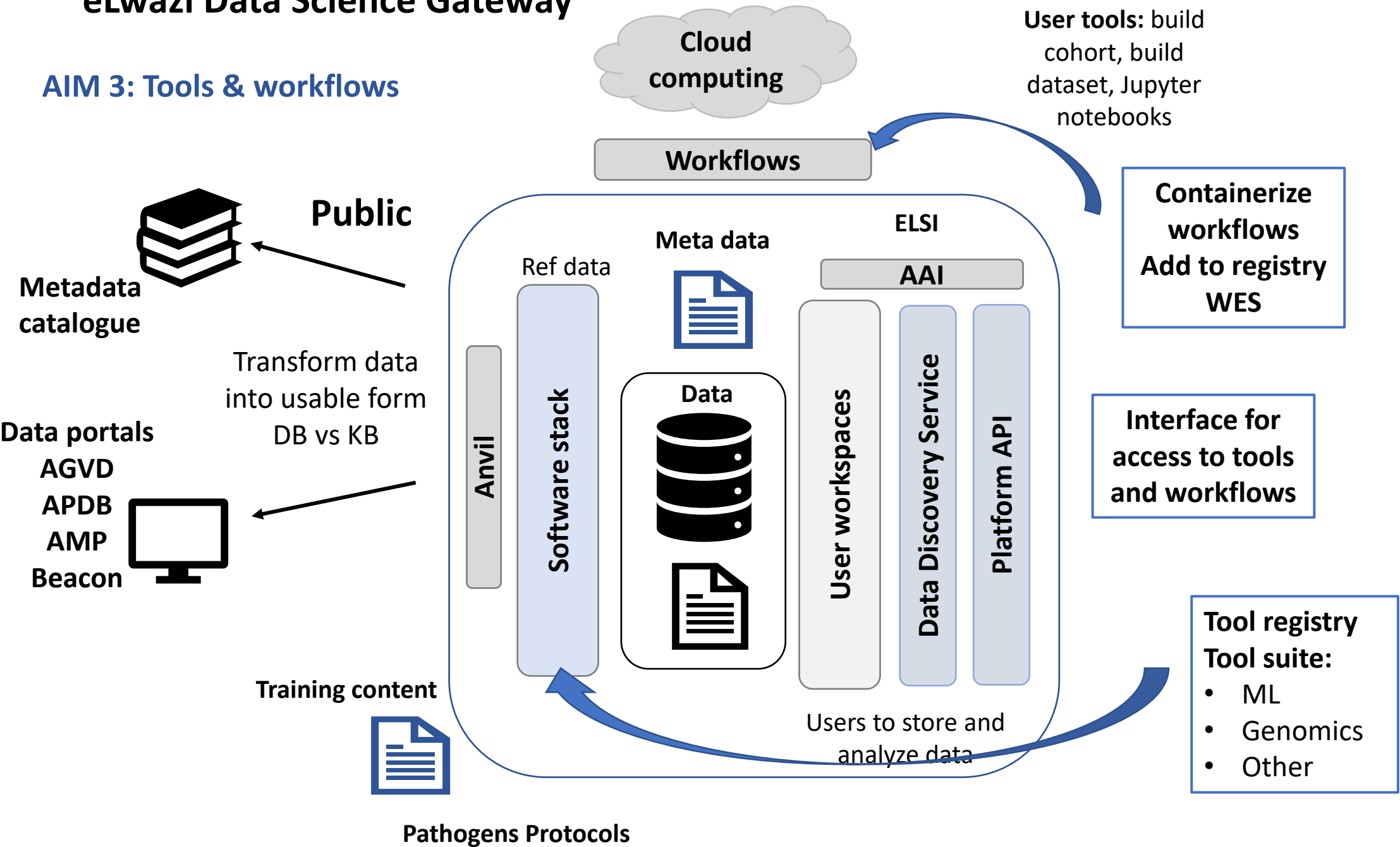


Aim 3. Tools and workflows

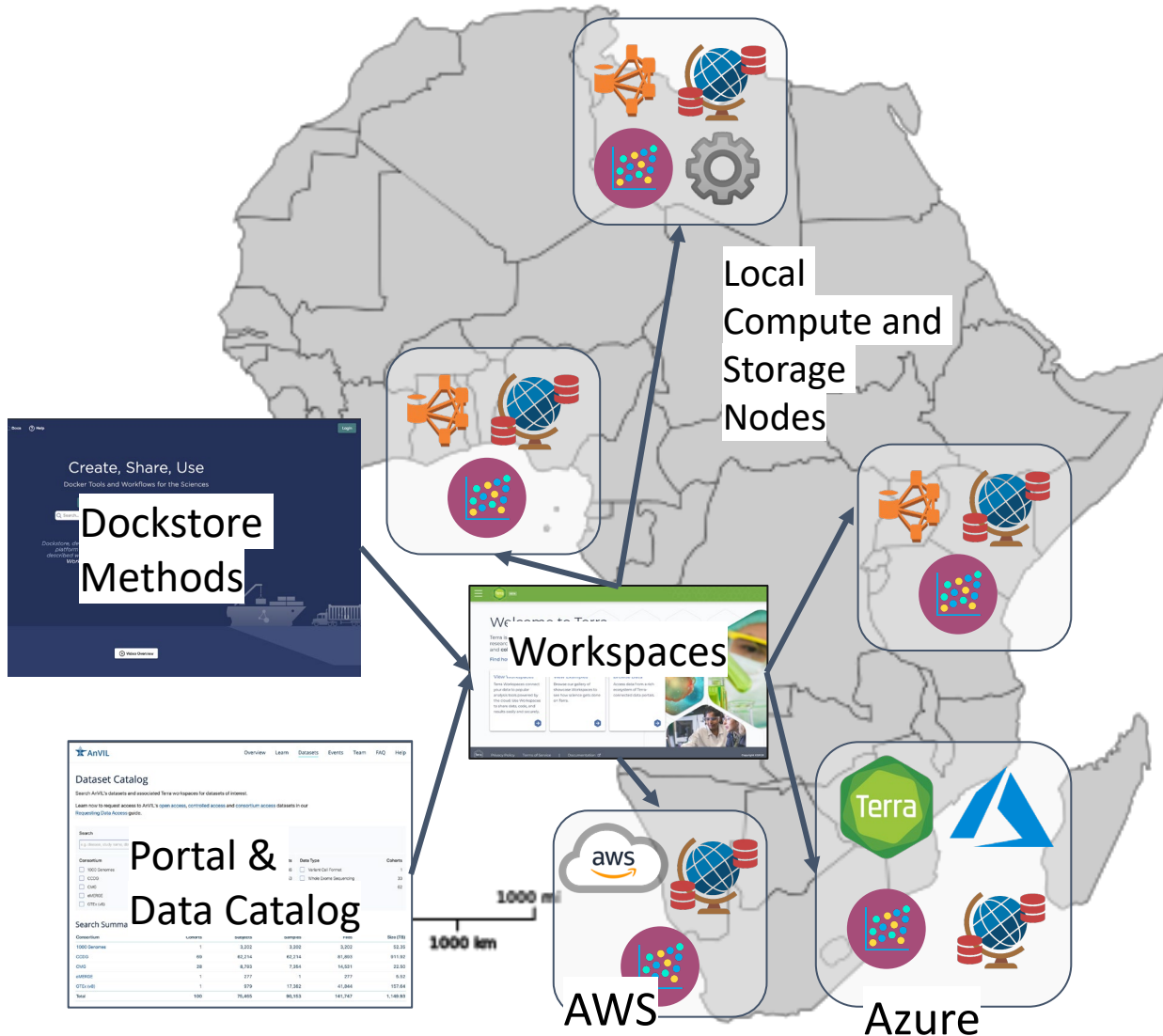
- Identify tools
- Installation on infrastructure
- Workflows
- Interface with Cloud providers
- Tool and workflow registries
- Galaxy
- Beginner and advanced access to tools

eLwazi Data Science Gateway

AIM 3: Tools & workflows



DS-I Africa ODSP hosting sites



Cloud & HPC Environments

-  AWS,  Azure, and  Local


 **DRS** for data access

 **WES** for compute

 **TRS** for workflow sharing



Workspaces for:

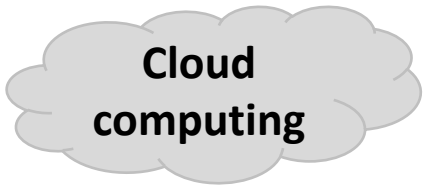
-  Data modeling and access
- Compute on data in various locations
- Interactively analyze via Notebooks
- Sharing and Collaborating

Aim 4. User support

- User support type
 - User group meetings
 - Use case support
 - Clinical data support
- Helpdesk
- Training
 - Curriculum standards
 - Professional training
 - Coordinated with training programs

eLwazi Data Science Gateway

AIM 4: User support



User tools: build cohort, build dataset, Jupyter notebooks

Workflows

Access for novice and advanced users

Public

Metadata catalogue

Transform data into usable form
DB vs KB

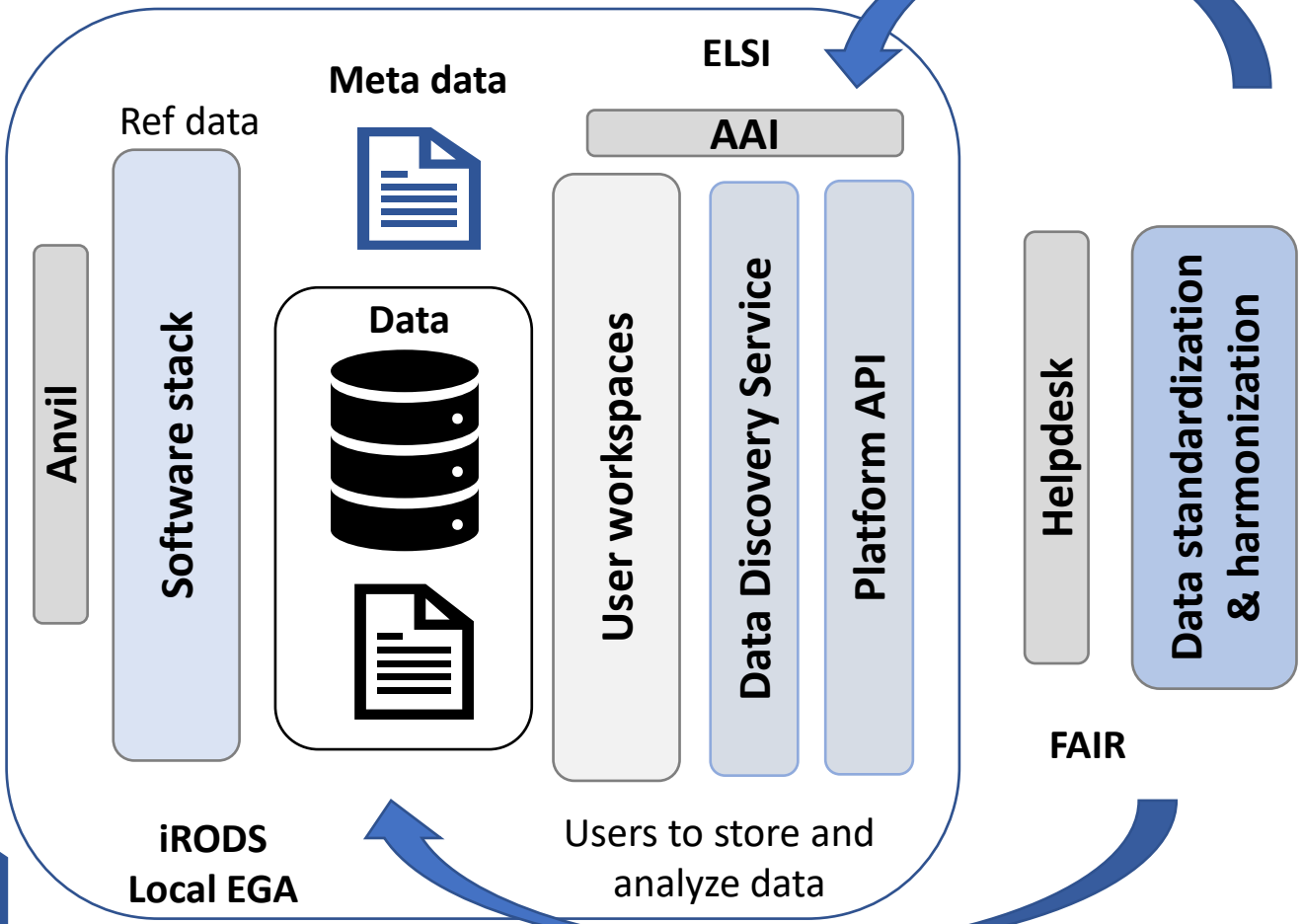
Data portals

- AGVD
- APDB
- AMP
- Beacon

Training Grants

Training content

Pathogens Protocols



Research hub

Research hub

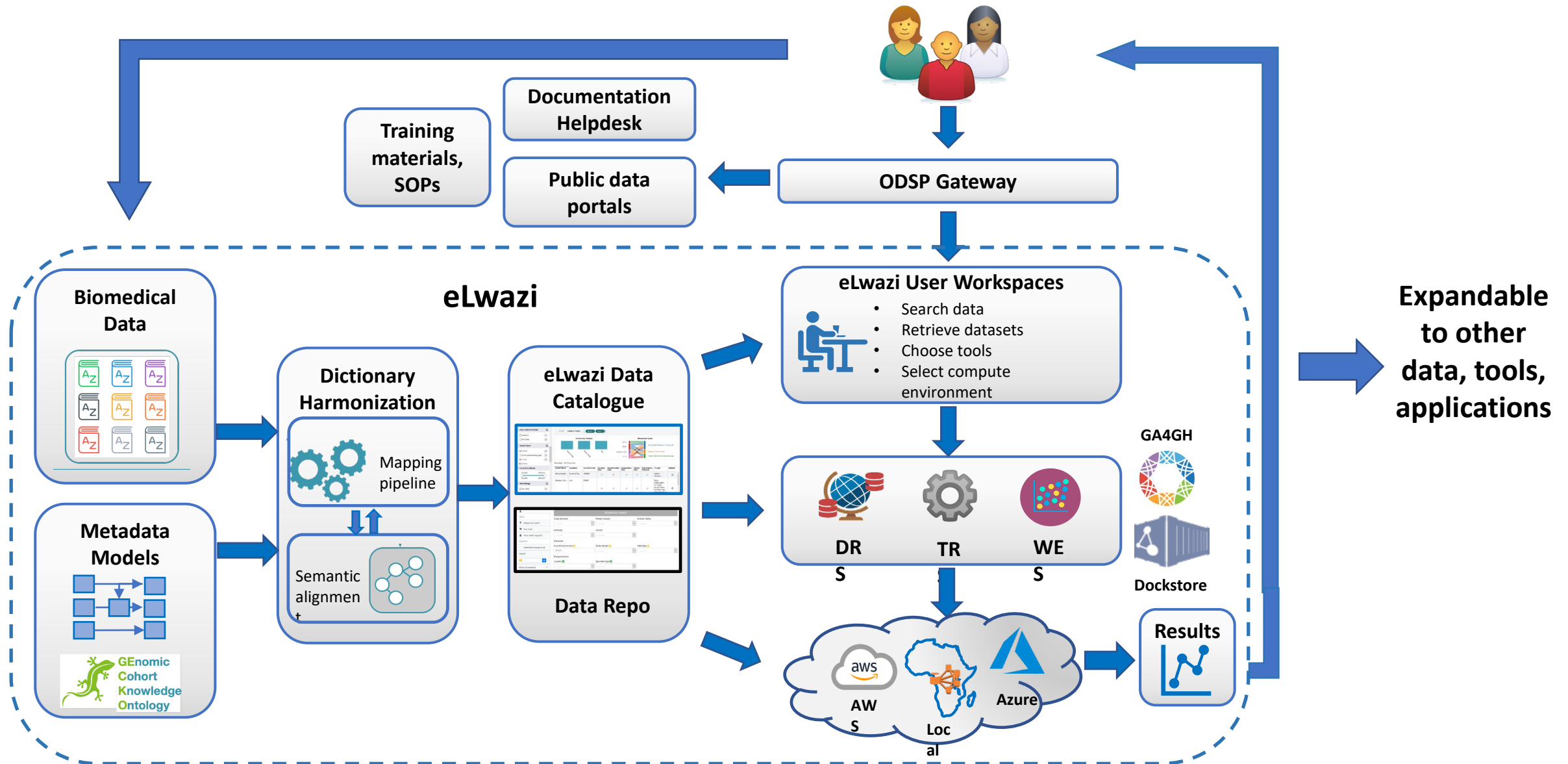
Research hub

REDCap support
PRDB

Aim 5. Monitoring, outreach, sustainability

- Monitor platform usage and impact
- Dissemination
- Develop cost model and sustainability plan
- Engagement with external stakeholders
 - NRENS
 - Cloud providers
 - Industry

Overview of the ODSP



Acknowledgements

Coordinating Centre PI: Dr Michelle Skelton

ODSP Partners

- University of Cape Town
- University of the Witwatersrand
- University of the Western Cape (Ilifu)
- University of Kwazulu-Natal -> Stellenbosch
- University of Mauritius
- Uganda Virus Research Institute
- University of Khartoum
- USTTB, University of Bamako
- Broad Institute
- EMBL European Bioinformatics Institute
- University of California, Santa Cruz
- University of Chicago

Funding: DS-I Africa program through the NIH Common Fund