



# **INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools**

January 2024 *Frontiers in Digital Health* 6:3

**Tathagata Bhattacharjee<sup>1\*</sup>, Sylvia Kiwuwa-Muyingo<sup>2\*</sup>, Chifundo Kanjala<sup>1,3</sup>, Molulaqhoa L. Maoyi<sup>4</sup>, David Amadi<sup>1</sup>, Michael Ochola<sup>2</sup>, Damazo Kadengye<sup>2,5</sup>, Arofan Gregory<sup>6</sup>, Agnes Kiragga<sup>2</sup>, Amelia Taylor<sup>7</sup>, Jay Greenfield<sup>5</sup>, Emma Slaymaker<sup>1</sup>, Jim Todd<sup>1</sup> and INSPIRE Network<sup>8</sup>**

<sup>1</sup>Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom,

<sup>2</sup>African Population and Health Research Center (APHRC), Nairobi, Kenya,

<sup>3</sup>UNICEF (Malawi), Lilongwe, Malawi,

<sup>4</sup>South African Population Research Infrastructure Network (SAPRIN), South African Medical Research Council, Durban, South Africa,

<sup>5</sup>Department of Economics and Statistics, Kabale University, Kabale, Uganda,

<sup>6</sup>Committee on Data of the International Science Council (CODATA), Paris, France,

<sup>7</sup>Malawi University of Business and Applied Sciences, Blantyre, Malawi,

<sup>8</sup>Implementation Network for Sharing Population Information from Research Entities (INSPIRE Network), Nairobi, Kenya

DOI: [10.3389/fdgth.2024.1329630](https://doi.org/10.3389/fdgth.2024.1329630)

# OBJECTIVE

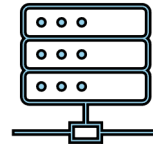
A pan-African, Findable, Accessible, Interoperable, and Reusable (FAIR) research architecture and infrastructure

Cloud-based Platform-as-a-Service (PaaS) and on-premises setup aimed to enhance the discovery, integration, and analysis of clinical, population-based surveys, and other health data sources.

# INSPIRE Databhub

## Problem

- 1 Population health data integration in LMICs
- 2 Generation of actionable insights
- 3 Informed policy and decision-making
- 4 Common platform for data analysis



## Solution

- 1 A pan-African platform
- 2 Harmonises longitudinal population health data to OMOP Standards
- 3 Addresses data integration challenges in LMICs
- 4 FAIR (Findable, Accessible, Interoperable, and Reusable)

# The Platform



INSPIRE datahub includes key components such as virtualization, cloud computing, and the OMOP Common Data Model (CDM).

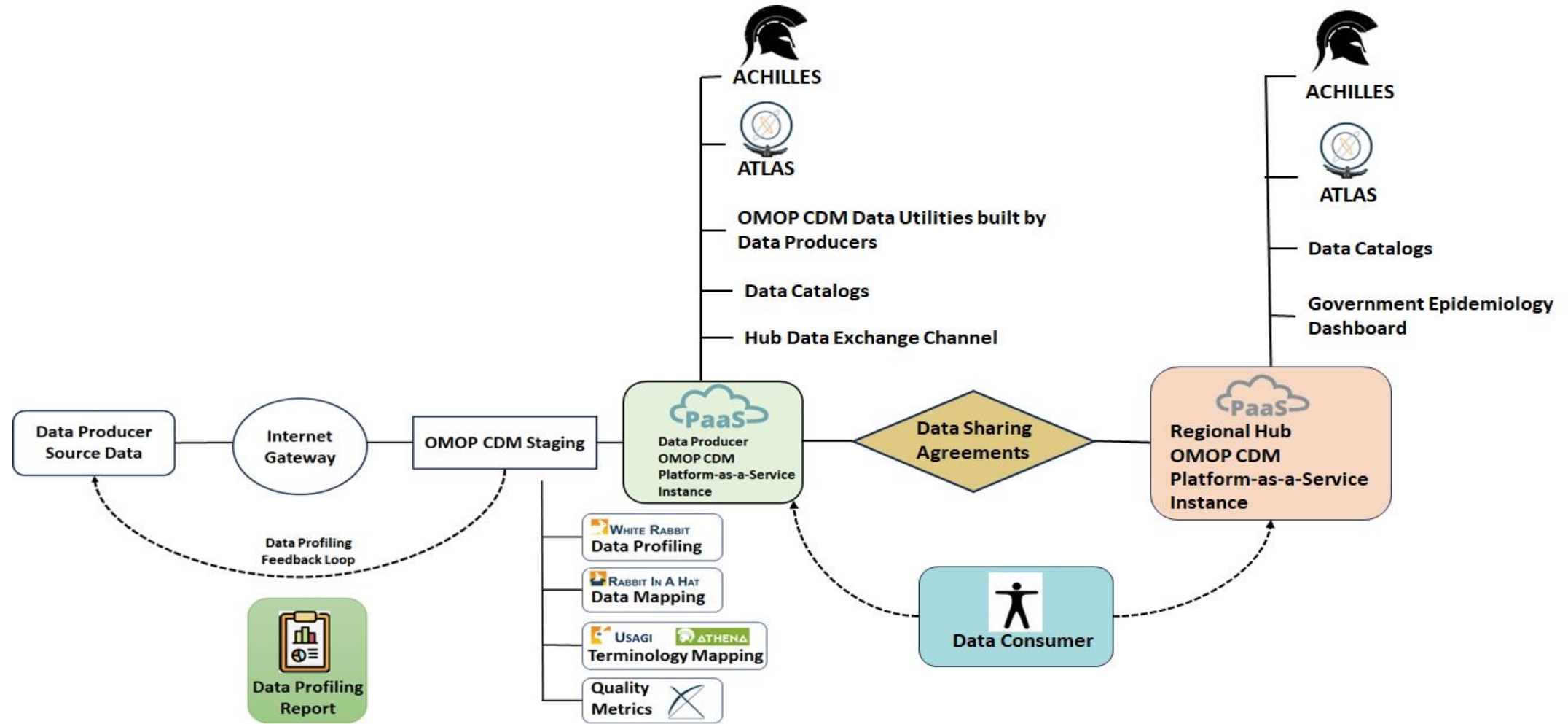


Employs data mapping and migration techniques to ensure seamless data flow from source to OMOP CDM

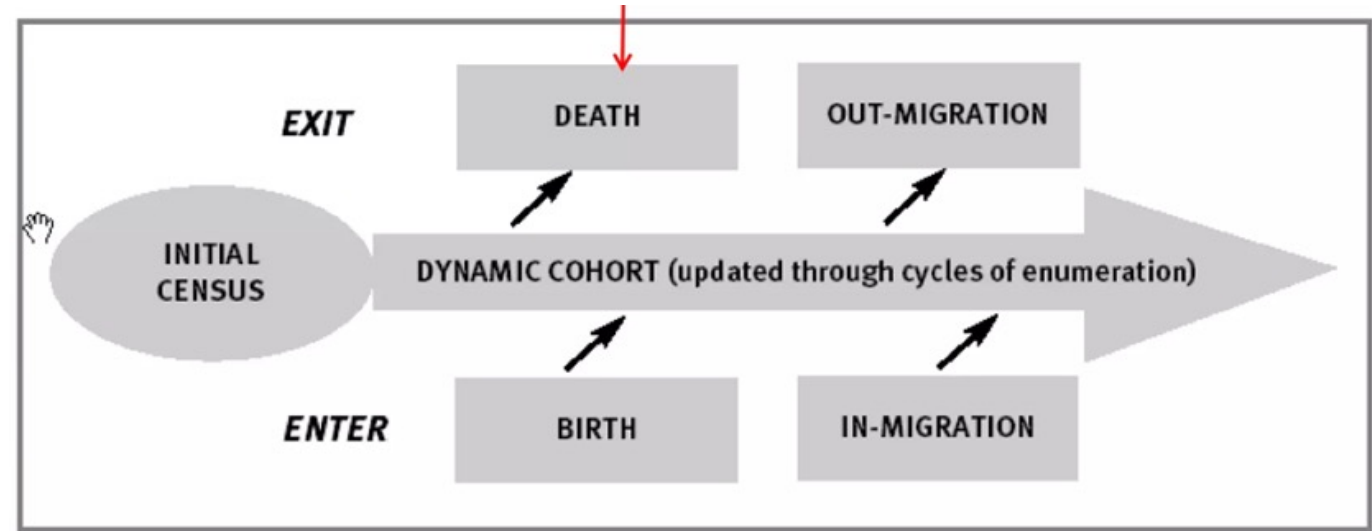


Promotes federated research network for collaborative research efforts

# INSPIRE datahub overall architecture

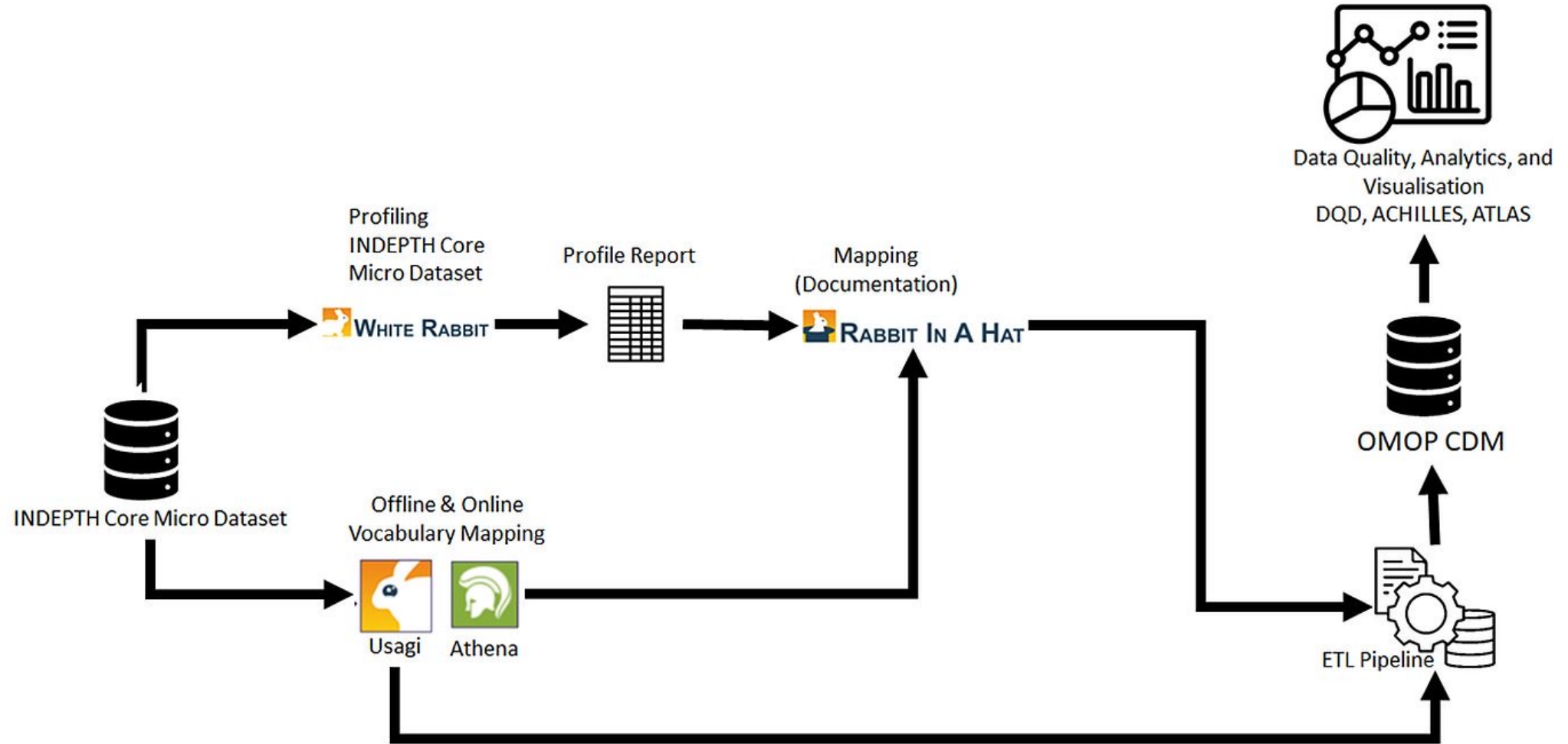


# Case-study



HDSS Site	Male	Female	Other	Total
Kenya – Nairobi	118,487	97,907	0	216,394
Tanzania Magu(Kisesa)	49,302	56,295	35	105,632
South Africa – AHRI	89,244	106,002	0	195,246
<b>Total</b>	<b>257,033</b>	<b>260,204</b>	<b>35</b>	<b>517,272</b>

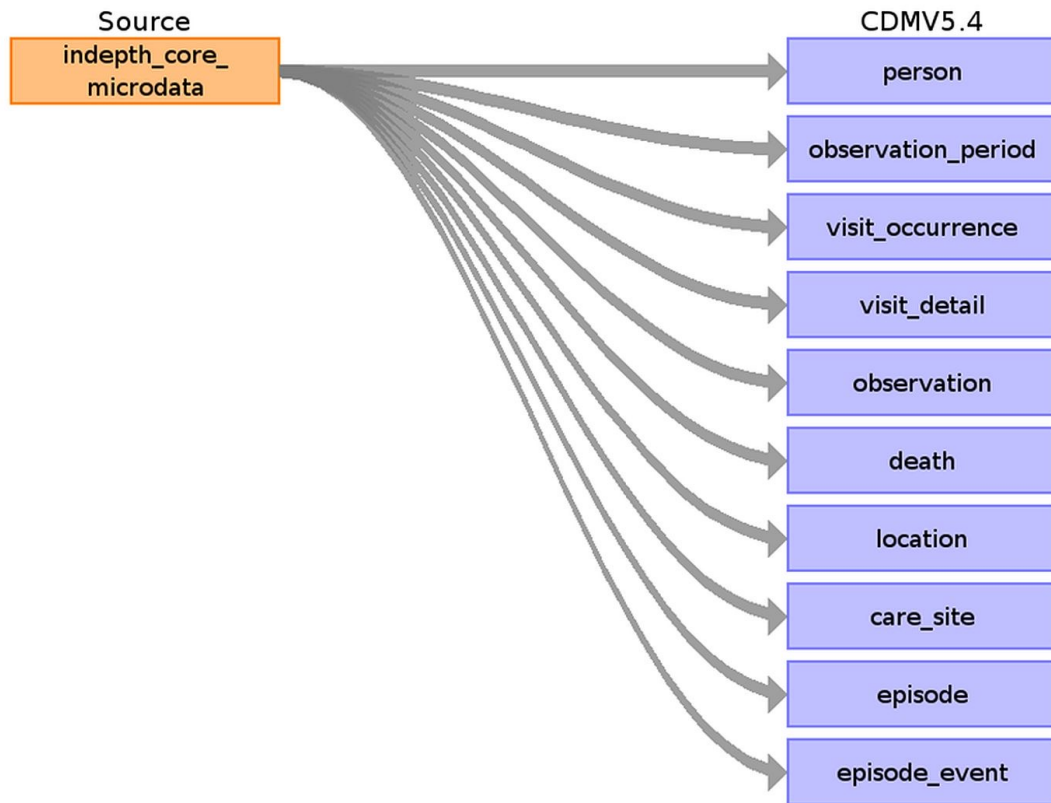
# Data flow from INDEPTH data repository to INSPIRE and to OMOP CDM



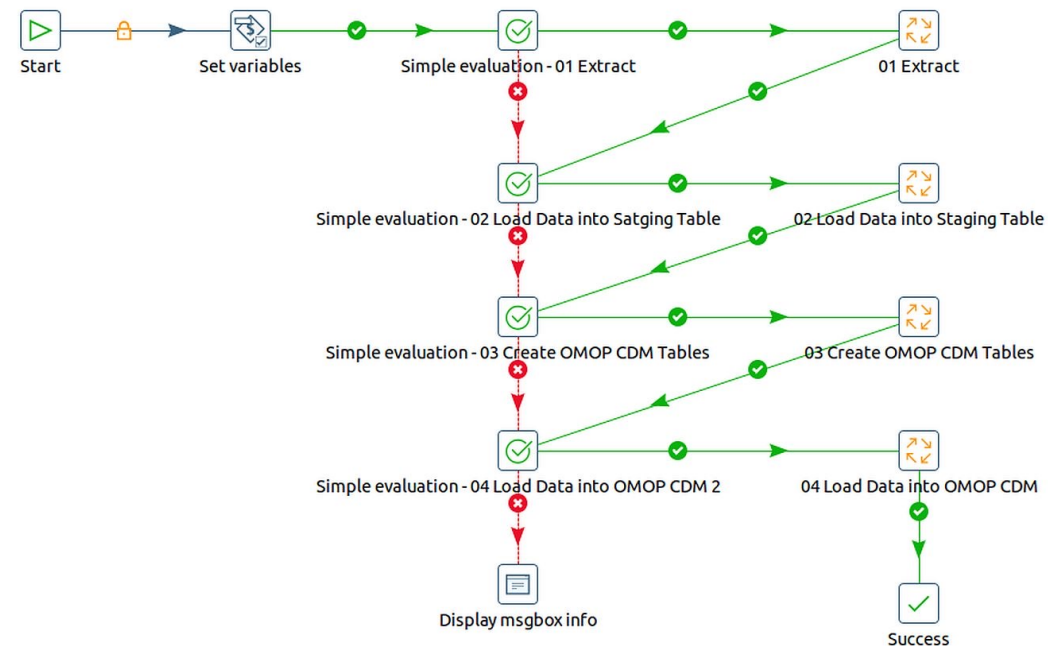
# Data Migration

Mapping from INDEPTH Core Micro Dataset to OMOP CDM

## Source Data Mapping Approach to CDMV5.4

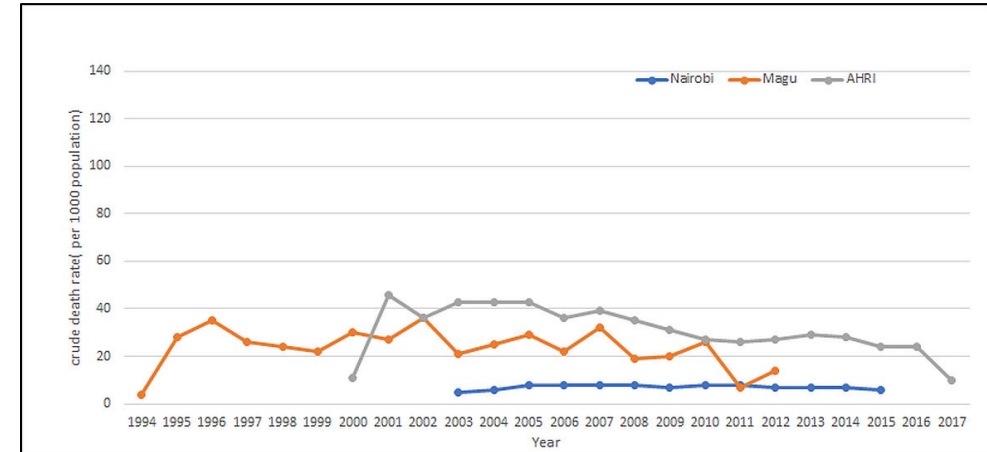
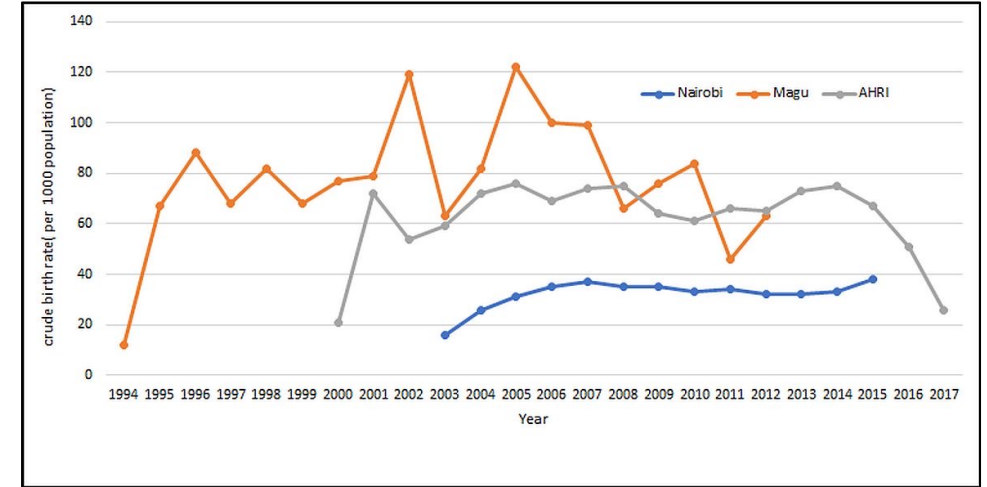
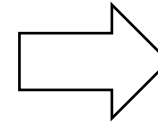
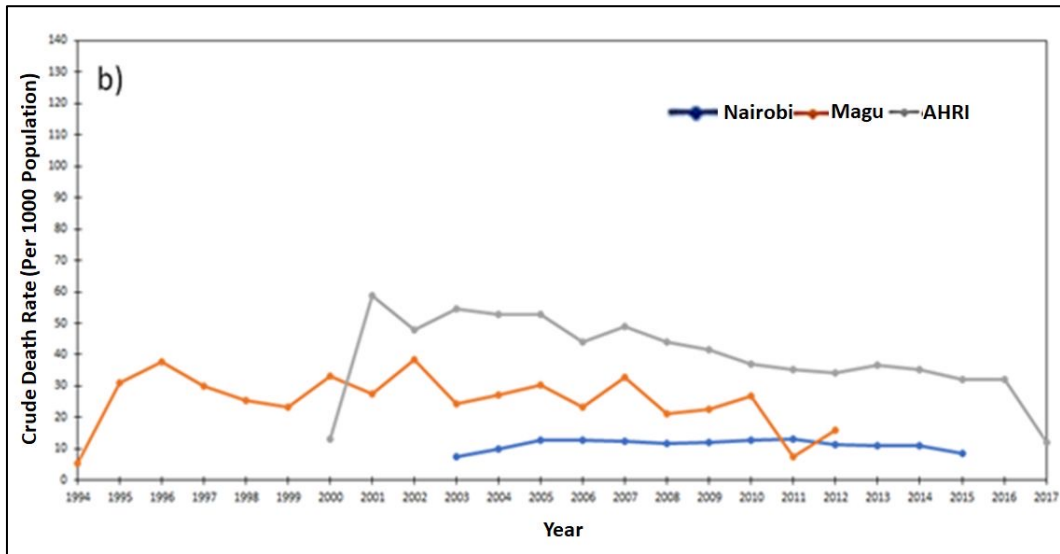
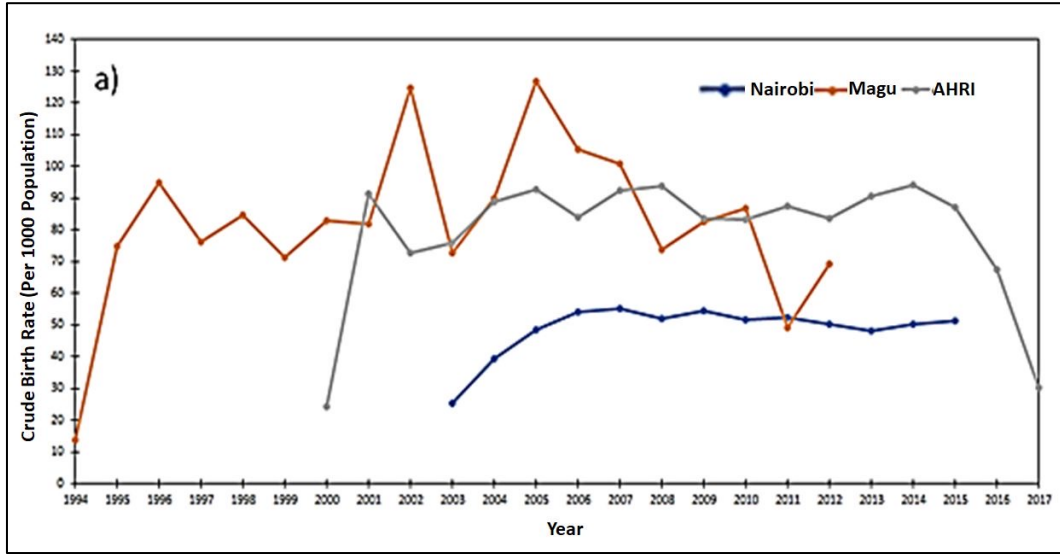


Implementation of Data Migration (ETL) using Pentaho Data Integration tool





# Before and After Migration



# Result

## Data Quality Dashboard

ATLAS

### DATA QUALITY ASSESSMENT

#### CDM INDEPTH CORE MICRO DATASET TO OMOP V1.0

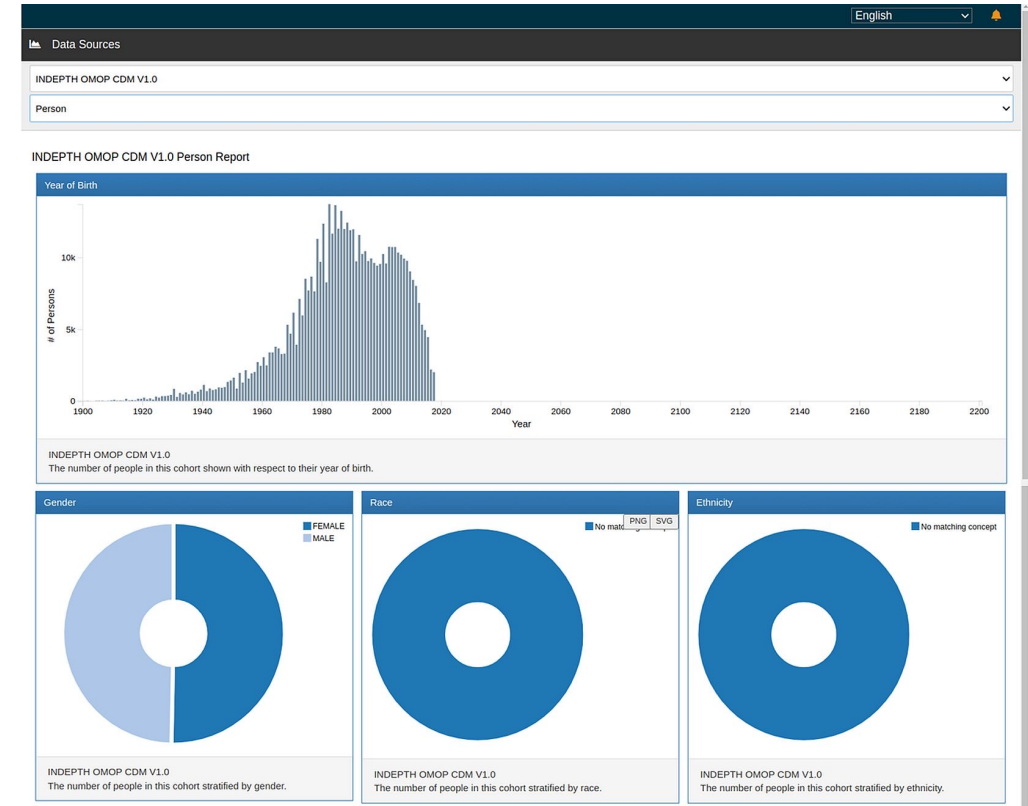
DataQualityDashboard Version: 2.4.0  
Results generated at 2023-10-16 19:28:30 in 7 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	2142	29	2171	99%	287	0	287	100%	2429	29	2458	99%
Conformance	689	39	728	95%	106	0	106	100%	795	39	834	95%
Completeness	378	18	396	95%	17	0	17	100%	395	18	413	96%
<b>Total</b>	<b>3209</b>	<b>86</b>	<b>3295</b>	<b>97%</b>	<b>410</b>	<b>0</b>	<b>410</b>	<b>100%</b>	<b>3619</b>	<b>86</b>	<b>3705</b>	<b>98%</b>

3048 out of 3619 passed checks are Not Applicable, due to empty tables or fields.  
40 out of 86 failed checks are SQL errors.  
Corrected pass percentage for NA and Errors: 93% (571/617).

#### Data Source Metadata

CDM INDEPTH Core Micro Dataset to OMOP V1.0  
INDEPTH Core Microdata  
Licensed to: OHDSI  
Source Released: 2023-10-15  
CDM Released: 2023-10-15  
CDM Version: 5.4



# THANK YOU

Tathagata Bhattacharjee

tathagata.bhattacharjee@lshtm.ac.uk

<https://www.lshtm.ac.uk> | <https://aphrc.org/inspire> | <https://www.inspiredata.network>