

data_dictionary

Christina Myalla

2026-02-13

Health Facility Data Dictionary

Overview

This document describes the standardized variables produced by the multi-country health facility ETL pipeline.

All country datasets are harmonized into a common structure to support analysis, mapping, and integration across countries.

Dataset Structure

Each row represents one health facility.

Variables

Variable Name	Type	Description	Example	Notes
facility_name	character	Official facility name	Mwananyamala Hospital	Standardized from country raw field
facility_code	character	Unique facility identifier	HF12345	May come from national registry
facility_type	character	Facility classification	Hospital, Clinic, Dispensary	Standardized labels
facility_ownership	character	Managing authority	Public, Private, Faith-based	Harmonized categories
latitude	numeric	Geographic latitude (WGS84)	-6.7924	May be missing if unavailable
longitude	numeric	Geographic longitude (WGS84)	39.2083	May be missing if unavailable
admin1	character	First administrative level	Dar es Salaam	Region / State / Province

Variable Name	Type	Description	Example	Notes
status	character	Operational status	Operating, Closed	Optional by country
valid_coords	logical	Coordinate inside country boundary	TRUE / FALSE	Derived spatial validation

Coordinate System

All spatial coordinates are stored in:

- CRS: WGS84
 - EPSG: 4326
-

Missing Values

Field	Meaning
NA latitude / longitude	Coordinates not available
NA status	Not provided by source country

Derived Fields

valid_coords

Indicates whether facility coordinates fall within national boundaries.

- TRUE → inside country polygon
 - FALSE → outside country polygon
 - NA → coordinates missing
-

Data Sources

Country-specific health facility registries or master facility lists. See `docs/country_notes/` for detailed metadata.

Versioning

Schema version: 1.0

Last updated: YYYY-MM-DD