

# transformation\_rules

Christina Myalla

2026-02-13

## Transformation Rules

### Purpose

Defines all data cleaning, harmonization, and validation rules applied during transformation.

---

### Column Standardization

Country-specific raw column names are mapped to global schema fields.

Example:

Raw Column	Standard Column
facility	facility_name
lat	latitude
lon	longitude

Mappings defined in `config/countries.yml`.

---

### Data Type Enforcement

Field	Rule
latitude	numeric
longitude	numeric
facility_name	character
facility_code	character

Invalid conversions become NA.

---

## Coordinate Handling

- Missing coordinates are retained
  - Non-numeric values converted to NA
  - Spatial validation applied
- 

## Spatial Validation

Coordinates checked against national boundary polygon.

Output:

Result	Meaning
TRUE	inside country
FALSE	outside
NA	missing coordinates

---

## Facility Ownership Harmonization

Ownership categories standardized where possible.

Example mapping:

Raw	Standard
Govt	Public
Private NGO	Private
FBO	Faith-based

---

## Facility Type Harmonization

Facility types standardized to controlled vocabulary.

Examples:

- Hospital
  - Health Centre
  - Clinic
  - Dispensary
-

## **Variable Selection**

Only core analytical variables retained:

- facility\_name
  - facility\_code
  - facility\_type
  - facility\_ownership
  - latitude
  - longitude
  - admin1
  - status
  - valid\_coords
- 

## **Error Handling**

Pipeline stops when required configuration fields are missing.

Optional fields are allowed and replaced with NA if absent.

---

## **Logging**

Processing messages generated for:

- missing configuration
- invalid coordinates
- dropped columns