



Data Release Notes

Sierra Leone National Health Facilities Dataset Version 01

July 2023

Abstract

This document outlines the methodological approach and data sources used to construct the Sierra Leone National Health Facilities Dataset Version 01 dataset. The dataset consists of health facility points with name, location (where available), facility type, functionality, and ownership in Sierra Leone. Limitations and terms of use are provided.

Dataset citation

Sierra Leone Ministry of Health and Sanitation (MoHS) and Center for International Earth Science Information Network (CIESIN), Columbia University. 2023. Sierra Leone National Health Facilities Dataset Version 01. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). <https://doi.org/10.7916/ag19-s334>. Accessed DAY MONTH YEAR

Terms of Use

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

The data and accompanying document are licensed under a Creative Commons Attribution 4.0 International License, CC BY 4.0 (<http://creativecommons.org/licenses/by/4.0>) and specified in legal code (<http://creativecommons.org/licenses/by/4.0/legalcode>).

Copyright 2023. Sierra Leone Ministry of Health and Sanitation (MoHS) and Trustees of Columbia University in the City of New York.

Contacts and data queries

The authors of this dataset appreciate feedback regarding the data, including suggestions, discovery of errors, difficulties in using the data, and format preferences. For dataset-related questions, please send an email to any of the following organizations:

GRID3: info@ciesin.columbia.edu

MoHS: dir.dppi@mohs.gov.sl

1. Introduction

This work was conducted in collaboration between the Directorate of Policy Planning and Information (DPPI) within Sierra Leone's Ministry of Health and Sanitation (MoHS) and the GRID3 programme hosted at CIESIN, Columbia University. Through engagement with DPPI, as well as non-government partners, CIESIN collected and consolidated existing sources of health facility data. Spatial data were matched to the ministry's master facility list (2020). Additional data collection was necessary and conducted by GRID3's partner, Premise, in coordination with the ministry. A data validation workshop was held in Bo, Sierra Leone, 28-30 July 2021, where MoHS District Medical Officers (DMOs), District Monitoring and Evaluation (M&E) Officers, Directors, Managers, and key stakeholders reviewed and provided feedback on the dataset. Outstanding data points were submitted by District M&E Officers, and incorporated into the dataset by June 2023.

This work has been undertaken as part of the Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) programme. The programme is funded by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office. GRID3 is implemented by the Flowminder Foundation, WorldPop at the University of Southampton, the United Nations Population Fund, and CIESIN.

2. Methodological approach

2.1 Input datasets

Dataset Name	Description	Source
Consolidated Facility List_ UNICEF DPPI, 2016	Spatial data from the 2016 health facility mapping exercise	Ministry of Health and Sanitation (MoHS) and United Nations Children's Fund (UNICEF)
Draft Master Facility List_SARA WHO DPPI 20170706, 2017	Spatial data from the 2017 Service Availability and Readiness Assessment (SARA)	Ministry of Health and Sanitation (MoHS)
Updated List of Facilities_DPPI Mar 2020	Tabular dataset of health facilities. Considered the most up-to-date master facility list.	Ministry of Health and Sanitation (MoHS)
FMFL with GPS Coordinates from MoHS, 2015	Spatial dataset of health facilities	Ministry of Health and Sanitation (MoHS)
Health Facilities in Sub-Saharan Africa, 2019	Spatial database of health facilities in various sub-Saharan countries	World Health Organization. (Updated February 2020). Health Facilities in sub-Saharan Africa. Humanitarian Data Exchange. Accessed November 16, 2020. https://data.humdata.org/dataset/health-facilities-in-sub-saharan-africa

Dataset Name	Description	Source
Health Facilities Lists - GPS Coordinates_03Mar2016, 2016	Spatial data from the National Emergency Medical Service (NEMS), formed in 2015	Ministry of Health and Sanitation (MoHS), National Emergency Medical Service (NEMS)
Health Facilities List - eidsr_covid19_dhmt_update_Jul2020	Spatial data of health facilities collected in the field using smartphones, as part of COVID19 support efforts. <i>Note: This dataset was used for validation purposes only.</i>	World Health Organization's Country Office in Sierra Leone

2.2 Methods

CIESIN performed an exploratory data analysis on each dataset to understand their respective features, missing data (if any), and typology. The Updated List of Facilities_DPPI from March 2020 file was determined to be the most comprehensive and recent. This was confirmed by the ministry and thus the dataset was considered the benchmark to complete or “add to”.

1. Data standardization and cleaning

- a. Basic data cleaning was performed on the attribute data in all input files:
 - i. Replaced dashes with spaces
 - ii. Created separate columns for facility name and facility type
 - iii. Fixed basic spelling errors in facility types
- b. Specifically in the SARA data, spatial cleaning of the “Lat” and “Long” fields was performed. The latitude and longitude were transposed for some data points, causing them to fall outside of Sierra Leone. These points were identified, and the latitude and longitude were flipped to correct the error.

2. Checking spatial quality

The availability of multiple independently collected datasets allowed confirmation of the spatial location and attribute data of most of the health facilities by following these steps:

- a. All datasets with spatial information were merged into one dataset.
- b. Health facility points within 100 meters of each other were made into a cluster.
- c. The health facility names within each cluster were evaluated to see if they agreed on one common health facility name.
- d. Clusters were scored based on the number of data sources that agreed on a name and place of a health facility.
- e. Clusters where two or more data sources agreed on a name and location were chosen over clusters with only a single instance.

3. Matching to the MFL benchmark

- a. Using the Levenshtein distance algorithm, geospatial data were matched to the tabular benchmark data (i.e. Updated List of Facilities_DPPI Mar 2020 data file) using a combination of the chiefdom and health facility name. All matches were manually inspected. In cases of uncertainty, the ministry was consulted.

4. *Quality Checks*

- a. Confirmed that matched points fell into the correct chiefdom boundary. Points were marked if there was a question about spelling or if a chiefdom name may have changed.
- b. Confirmed that matched points fell within a settlement extent. If a point fell outside of a settlement extent, the point was visually inspected. If there was no apparent health facility at that location, the point was removed from the match.
- c. Confirmed that the type of the facility matched the type listed in the benchmark file. If there was a question about the type of facility, the point was marked for consultation with the ministry.

Health facilities that were matched with at least two independently collected data sources (i.e the 2017 SARA and 2020 DPPI list) were considered to be valid and have confirmed locations. The other facilities with only one spatial data source matched were considered “unvalidated”; the facilities with no spatial match were considered “missing”.

The unvalidated and missing health facilities were flagged for further field validation, either to fill in the gaps or to confirm the location of a facility. GRID3 partner, Premise, conducted data collection through its mobile platform and network of local contributors. The name and, when available, additional geographic information (such as chiefdom and name of locality where the health facility was thought to be located) were shared with Premise and integrated into the data collection tasks. The spatial data collected through this process are incorporated in the dataset and considered in the MFL alpha version released in July 2021.

2.3 Validation workshop and version history

Following the data collection exercise, validation workshops were conducted after the release of the MFL alpha version in July of 2021 and on the final version in May of 2023, in order to review and collect feedback from main stakeholders. District-level MoHS staff systematically reviewed each health facility record. Corrections were made to the data attributes (including the facility’s location) with the aid of Excel documents and maps. Additional health facilities were identified during the workshops. These records were shared for additional field validation. Not all health facilities added during the workshop were validated with a GPS location. These facilities are located based on the city center or settlement name in which they reside.

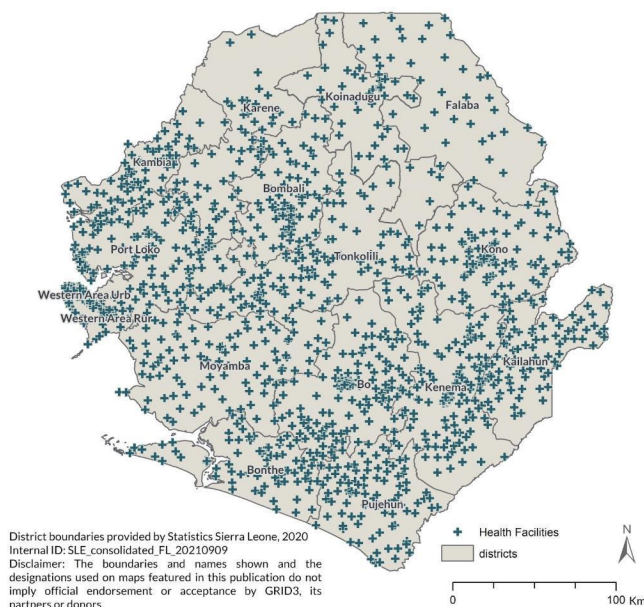
2.4 Dataset description

The data is available in both spatial and tabular file formats. A map of data is shown below.

Extent: Sierra Leone: Admin Level 0 Boundaries.

Coordinate system: GCS WGS 1984

Sierra Leone National Health Facilities, Version 01



2.5 Dataset codebook

Field Name	Field Type	Description
OBJECTID	OBJECT ID	Default Esri system-managed value that uniquely identifies a record or feature.
Shape	Geometry	Default Esri system-managed value that uniquely identifies a record or feature type
district	Character	District name, source: Statistics Sierra Leone (2020).
settlement	Character	Settlement name
facility	Character	Health facility name which includes the facility type
fac_name	Character	Health facility name, short name
type	Character	Health facility type
ownership	Character	Ownership of the health facility: Government, Private, Faith-based, or NGO
functional	Character	Functional status of the health facility: Functional or Not Functional
long	Numeric	Longitude, in decimal degrees
lat	Numeric	Latitude, in decimal degrees
source	Character	Source of health facility name (Note: "DPPI_ND" where the ND means "No Date")
source_geo	Character	Source of the health facility location: GPS, City Center, or Digitized
dhis2_id	Character	Unique identifier from DHIS2, when a match was

Field Name	Field Type	Description
		available. When no match was available, "No Data" was added.
date	Character	Date when location data was obtained
Comment		Comments

3. Known data limitations and disclaimer

Although much effort was put into collecting the locations, 30 locations are still unvalidated and 63 are still without geospatial information. The majority of the unvalidated/ non-geolocated health facilities are in areas known to be difficult to access. In addition, the dataset may contain a number of errors and/or omissions. Suggested edits and feedback are welcome, and will be reviewed and incorporated into subsequent versions of the dataset.

Sierra Leone's MoHS, CIESIN at Columbia University, and the GRID3 partners follow procedures designed to ensure that data disseminated by the project are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact authors at info@ciesin.columbia.edu or MoHS at dir.dppi@mohs.gov.sl. Sierra Leone's MoHS, CIESIN at Columbia University, and their sponsors do not guarantee the accuracy, reliability, or completeness of any data provided. We provide these data without warranty of any kind whatsoever, either expressed or implied, and shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided.

4. Acknowledgments

The MoHS in Sierra Leone and GRID3 would like to thank the United Nations Children's Emergency Fund (UNICEF), the National Emergency Medical Service (NEMS) in Sierra Leone, Partners in Health (PIH) and World Health Organisation (WHO) for their contributions to this dataset. Funding for the development and dissemination of this dataset was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office.