

Data Release Statement
GRID3 DRC Haut-Lomami and Tanganyika Health Facilities,
Version 01 (Beta)

31 August 2020

Abstract

This document outlines the methodological approach and data sources used to construct the *GRID3 DRC Haut-Lomami and Tanganyika Health Facilities, Version 01 (Beta)* dataset. The dataset consists of health facility points with name, location, health zone and health area attributes in the provinces of Haut-Lomami and Tanganyika in the Democratic Republic of the Congo (DRC). Limitations and use constraints are provided.

Dataset citation:

Ministère de la Santé Publique, République Démocratique du Congo and Center for International Earth Science Information Network (CIESIN), Columbia University. 2020. *GRID3 DRC Haut-Lomami and Tanganyika Health Facilities, Version 01 (Beta)*. Palisades NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). DOI: <https://doi.org/10.7916/d8-tnx3-8z95>. Accessed DAY MONTH YEAR

Data Use Constraints:

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.

Copyright 2020. Ministère de la Santé Publique en Kinshasa, République Démocratique du Congo and The Trustees of Columbia University in the City of New York.

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

Contacts and Data Queries

GRID3 appreciates feedback regarding this dataset, such as suggestions, discovery of errors, difficulties in using the data, and format preferences.

Please contact: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3), data.queries@grid3.org



Contents

1.	Introduction.....	3
2.	Methodological Approach.....	4
3.	Dataset Description	7
4.	Known Data Limitations.....	8
5.	Disclaimer	9
6.	Acknowledgements	9
7.	References.....	10
8.	Appendix.....	11

1. Introduction

The *GRID3 DRC Haut-Lomami and Tanganyika Health Facilities, Version 01 (Beta)* dataset consists of health facility points with name, location, health zone and health area in the provinces of Haut-Lomami and Tanganyika in the Democratic Republic of the Congo (DRC). The dataset is one of three included in the *GRID3 DRC Haut-Lomami and Tanganyika, Version 01 (Beta)* data release. The other two datasets that are part of the release are *GRID3 DRC Haut-Lomami and Tanganyika Settlements, Version 01 (Beta)* and *GRID3 DRC Haut-Lomami and Tanganyika Health Catchment Area Boundaries, Version 01 (Beta)*.

This work was accomplished through engagement by the Center for International Earth Science Information Network (CIESIN) with the mandated authorities in the Ministry of Health in support of data collection and development for immunization planning. Local healthcare workers were directly involved in the mapping of the health catchment area boundaries at participatory events coordinated with in-country provincial coordinators and Geographic Information System technicians, and in the collection of data in the field from July to September 2019.

This work has been undertaken as part of the Geo-referenced Infrastructure and Demographic Data for Development (GRID3) programme in the Democratic Republic of the Congo (DRC). The programme is funded by the Bill & Melinda Gates Foundation and United Kingdom's Foreign, Commonwealth and Development Office. It is implemented by WorldPop/Flowminder (WPFM) and the United Nations Population Fund (UNFPA), and coordinated by the Center for International Earth Science Information Network (CIESIN) of Columbia University.

Input data

Table 1. *Input data layers*

Data	Source(s)
Baseline health facility data: compiled from various sources	(WHO, 2019); (Programme de Santé Intégré de l'USAID en République Démocratique du Congo (PROSANI USAID), 2019)
Health facility data collected in the field as part of the Ministry of Health's monitoring	(Acasus and Programme Elargi de Vaccination (PEV), 2020)
Health facility data collected in the field (raw data)	(Ministère de la Santé Publique, République Démocratique du Congo and Center for International Earth Science Information Network (CIESIN), 2020)

Input Health Facility Data Layer

The project took advantage of available health facility point data from various sources (Table 1). CIESIN collected, assessed, cleaned and compiled the available data into a single preliminary baseline health facility layer. During participatory mapping meetings, points in the baseline layer were validated and edited by mappers with input from local healthcare workers.

2. Methodological Approach

Summary

With the support of provincial and national health authorities, local healthcare workers and Congolese GIS specialists engaged in a participatory mapping process in Haut-Lomami and Tanganyika during the months of July, August and September 2019. All reference data layers described in Table 1 were used during participatory mapping meetings where at the direction of the healthcare workers, GIS technicians (or GIS mappers) edited the baseline health facility data. This included correcting misspellings, deleting erroneous points and editing associated attribute information. Following the participatory mapping meetings, additional health facility data were collected in the field and processed by the Center for International Earth Science Information Network (CIESIN).

Details

Phase 1: Consolidation and Cleaning of Existing Health Facilities Data

Desk research was conducted by CIESIN to identify and acquire existing open source health facilities datasets (Table 1) for Haut-Lomami and Tanganyika Provinces. Data from all sources underwent a series of quality assurance and quality control (QA/QC) processes which included:

1. Transforming the data to a common health facility schema
2. Cleaning health facility names, e.g., adding/removing spaces, handling special characters, etc.
3. Standardizing health facility type attribute values
4. Removing spatially proximate duplicate points with the same name

Phase 2: Participatory Mapping Meetings

Provincial coordinators helped organize teams of GIS mappers and local healthcare workers at participatory mapping meetings held at each health zone's headquarters in Haut-Lomami and Tanganyika provinces. A health zone is an administrative health unit that contains several health areas. Each healthcare worker at the meeting was

responsible for providing health services to one health area. The purpose of these 3-day meetings was to bring together the empirical knowledge of local health personnel and the technical knowledge of GIS mappers to edit the preliminary settlement and health facility baseline layers and produce a set of preliminary health catchment area boundaries. For the health facility layer, the process included:

1. Visually inspecting each health facility point
2. Flagging erroneous or duplicate points for deletion
3. Moving misplaced points to the correct location
4. Correcting the names and/or spelling of health facilities
5. Editing or adding corresponding attribute data (e.g., health area, facility type, etc.)

In addition to validating the existing data, points were added when a health facility present in the District Health Information Software (DHIS2) was missing. The validated layer was then sent back to CIESIN to undergo quality checks including:

1. Checking for geometry errors
2. Checking for completeness of attribute information
3. Checking points flagged for deletion and removing them
4. Checking the data against the master list of health facilities extracted from the DHIS2 for completeness of coverage
5. Correcting unique identifiers to match the DHIS2 identifiers

Phase 3: Field Data Collection

As part of the participatory meetings, local healthcare workers were trained to collect settlement place names, health facility locations, and points of interest (POI) (e.g., schools, rivers, health boundary limits, etc.) using an Open Data Kit (ODK)-based data collection app preloaded on tablets. Following the meetings, each healthcare worker returned to their health area to collect data using the tablets. The goals of the field data collection effort were to add missing settlements and health facilities identified during the participatory mapping meetings and to resolve ambiguities such as the correct name of a village. The GIS mappers and provincial coordinators then reviewed and cleaned the data. The data were then sent back to CIESIN for additional quality checks following the process below:

1. Normalising “type” field and cleaning of text attributes
2. Removing duplicate points (points with the same name located in a settlement extent for example)
3. Checking the data against the master list of health facilities for completeness of coverage
4. Removing pharmacies, unless those facilities were in the master list of health facilities

5. Removing points more than 100 metres from a settlement extent with no visible structures in imagery

Phase 4: Integrating existing data with data collected in the field

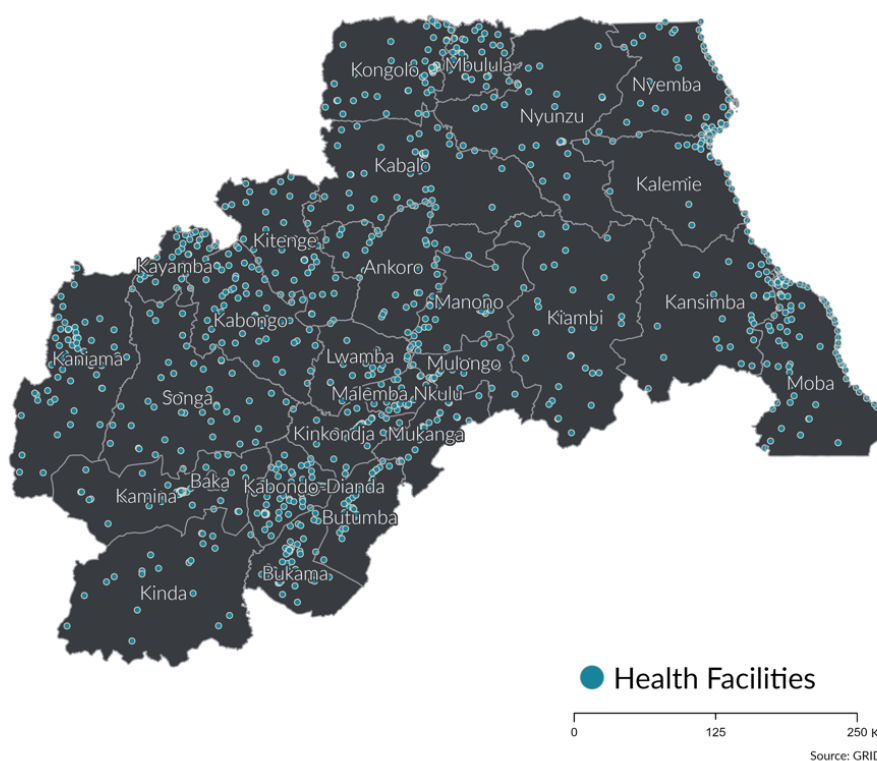
The data collected in the field proved to be highly accurate and nearly complete when compared to the master list of health facilities. In contrast, the quality of the baseline health facility data varied significantly by health zone as mappers reported insufficient time during the participatory mapping meetings to thoroughly validate and edit this layer. For these reasons, the following process was followed:

1. Data collected in the field were retained
2. Using data collected by Acasus, health facilities missing from the data collected in the field were appended if they existed in the master list of health facilities
3. Where gaps persisted, data from the baseline layer were appended
4. Health facility layer was checked against the master list for completeness and to ensure correct corresponding unique identifiers

3. Dataset Description

GRID3 DRC Haut-Lomami and Tanganyika Health Facilities Version 01 (Beta) dataset consists of one layer: health facilities point data. The data are available for download in shapefile or Esri file geodatabase formats packaged in zip files. See Appendix for a summary of the features in this dataset.

Health Facilities Version 01 (Beta)



Extent: Democratic Republic of the Congo: Haut-Lomami and Tanganyika Provinces

North: -4.97517681106092

West: 23.7154293063175

East: 30.5882263186897

South: -9.91798019413062

Coordinate system: GCS WGS 1984

Codebook

Field Name	Field Description
OBJECTID	Unique sequential numeric identifier maintained by the database.
province	The name of the province within which the health facility is located
zs_uid	Alphanumeric health zone (zone de santé) identifier as defined by the DHIS2 (Pyramide avec code INS)
zone_sante	The name of the health zone (zone de santé) within which the health facility area is located
as_uid	Alphanumeric health area (aire de santé) identifier as defined by the DHIS2 (Pyramide avec code INS)
aire_sante	The name of the health area (aire de santé) within which the health facility is located
fosa_uid	Health facility unique identifier as defined by the DHIS2 (Pyramide avec code INS) from DHIS2
fosa_nom	Health facility name
fosa_nom_2	Abbreviated health facility name
as_nom_alt	Alternative health facility name
type_abr	Abbreviated health facility type
lat	Latitude
lon	Longitude

4. Known Data Limitations

The spatial accuracy of the health facility data is dependent on both the accuracy of the point data collected in the field as well as on the correctness of the edits made to the layer compiled from existing health facilities at the participatory mapping meetings. In general, it was assumed that the field collected data were more accurate than the previously compiled health facility point data.

The scope of this project included fieldwork and validation for the health areas of 20 health zones in Haut-Lomami and Tanganyika Provinces. Existing data collected by the University of California Los Angeles in 7 of the health zones¹ (Lwamba, Mulongo, Mukanga, Malemba Nkula, Kikondja, Butumba and Manono) was incorporated into

¹ These data were collected by UCLA as part of the GIS layer used for the polio response in early 2018. Participatory mapping and high-resolution satellite imagery were used to delineate the boundaries.

the final product but has not been directly validated with local health teams under GRID3 supervision.

Known issues

A comparison of the health facility data set with the master list of health facilities extracted from DHIS2 indicates:

- One hundred and thirteen health facilities present in the DHIS2 master list do not have a corresponding record in this data set: 103 in Haut-Lomami and 10 in Tanganyika

5. Disclaimer

The data is provided "as-is". These data are part of ongoing work and are not guaranteed to be accurate and clean. If users encounter apparent errors or misstatements in the data, they should contact GRID3 at data.queries@grid3.org.

CIESIN, Columbia University, and their sponsors do not guarantee the accuracy, reliability, or completeness of any data provided. We provide this 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.

6. Acknowledgements

GRID3 thanks the following institutions that provided input data and/or assistance with data production.

Ministère de la Santé Publique, DRC
 Division du Système National d'Informations. Sanitaires (DSNIS), DRC
 Programme Elargi de Vaccination (PEV), DRC
 Division Provinciale de la Santé (DPS), DRC
 Bureau Central du Recensement (BCR), DRC
 Référentiel Géographique Commun (RGC), DRC
 The International Organization for Migration (IOM), DRC
 Programme de Santé Intégré de l'USAID en République Démocratique du Congo (PROSANI USAID), DRC
 Open Street Map (OSM), DRC
 Geospatial Evaluation and Observation Lab (geoLab), College of William & Mary, USA
 Acasus, Switzerland
 CartONG, France
 Centers for Disease Control and Prevention (CDC), USA
 DigitalGlobe, USA

Initiative Régionale de Documentation et d'Accompagnement Communautaire au Développement (IDRAC Sarl), DRC
 Médecins Sans Frontières (MSF)
 Novel-T, Switzerland
 Ministère de l'Environnement et Développement Durable (MEDD), DRC
 United Nations Children Fund (UNICEF), DRC
 United Nations Development Programme (UNDP)
 United Nations Office for Project Services (UNOPS)
 United Nations Organization Stabilization Mission (MONUC)
 United Nations Office for the Coordination of Humanitarian Affairs (OCHA)
 University of California, Los Angeles (UCLA) DRC Health Research and Training Program, USA
 World Health Organization (WHO)
 The consortium of the Bill & Melinda Gates Foundation and partners: PATH, VillageReach, Global Good, SANRU, Caritas, the International Medical Corps (IMC)
 World Resources Institute (WRI)

Funding for the development and dissemination of this dataset was provided by the Bill & Melinda Gates Foundation and United Kingdom's Foreign, Commonwealth and Development Office.

7. References

Center for International Earth Science Information Network (CIESIN), Columbia University and Novel-T. 2020. GRID3 Democratic Republic of the Congo Settlement Extents Version 01, Alpha. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). Source of building Footprints "Ecopia Vector Maps Powered by Maxar Satellite Imagery"© 2020. DOI: <https://doi.org/10.7916/d8-cpry-wv37>. Accessed DAY MONTH YEAR

Esri. "Bing Maps" [basemap]. Varying scale. "Bing Maps Aerial". Various dates. <https://www.arcgis.com/home/item.html?id=25c3f49d4ce3451e8a7f5b5aebccab48>. (December 2018-July 2019). Imagery (c) 2019 Microsoft Corporation and its data suppliers.

Esri. "World Imagery" [basemap]. Varying scale. "World Imagery Map". Various dates. <https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9>. (December 2018-July 2019).

Bureau Central du Recensement (BCR), République Démocratique du Congo and Center for International Earth Science Information Network (CIESIN), Columbia University. 2020. GRID3 DRC Haut-Lomami and Tanganyika Settlements, Version 01 (Beta). Palisades NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). DOI: <https://doi.org/10.7916/d8-zysy-qk80>. Accessed DAY MONTH YEAR

Ministère de la Santé Publique, République Démocratique du Congo and Center for International Earth Science Information Network (CIESIN), Columbia University. 2020. GRID3 DRC Haut-Lomami and Tanganyika Health Catchment Area Boundaries Version 01 (Beta). Palisades NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). DOI: <https://doi.org/10.7916/d8-v8vq-0p13>. Accessed DAY MONTH YEAR

8. Appendix

Distribution of health facilities by health zone and province

Province	Health Zone	Count of Health Facilities
Haut-Lomami	Baka	6
	Bukama	68
	Butumba	33
	Kabondo Dianda	4
	Kabondo-Dianda	64
	Kabongo	67
	Kamina	71
	Kaniama	64
	Kayamba	32
	Kinda	22
	Kinkondja	42
	Kitenge	43
	Lwamba	23
	Malemba Nkulu	27
	Mukanga	22
Mulongo	24	

Appendix (cont.)

Province	Health Zone	Count of Health Facilities
Tanganyika	Ankoro	34
	Kabalo	50
	Kalemie	122
	Kansimba	35
	Kiambi	32
	Kongolo	90
	Manono	29
	Mbulula	41
	Moba	162
	Nyemba	131
	Nyunzu	38

Note: Four health facilities are spatially located outside of health area boundaries but they are counted in the table.