COLUMBIA CLIMATE SCHOOL
CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK

Data Release Notes

Name of the dataset	GRID3 COD - Settlement Names v3.0
Name of the file	*GRID3_COD_settlement_names_v3_0.gpkg*
Date of data release	October 21, 2024
File format	OGC Geopackage
Dataset version	3.0
Abstract	This document outlines the methodology and data sources used during the production of the GRID3 COD - Settlement Names v3.0 dataset. The dataset consists of settlement points with name and attributes for thirteen provinces in the Democratic Republic of the Congo (COD). Limitations and use constraints are provided. This operational dataset has not been fully validated by government officials or ministries. The current version supersedes the GRID3 COD - Settlement Names v2.0. The following changes were made: ● Tshopo and Mongala Provinces have been added.
Dataset citation	Center for International Earth Science Information Network (CIESIN), Columbia University and Ministère de la Santé Publique, Hygiène et Prévention, Democratic Republic of the Congo, 2024. GRID3 COD - Settlement Names v3.0. New York: GRID3. https://doi.org/10.7916/y9hf-gm15. 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. Copyright 2024. The Trustees of Columbia University in the City of New York.
Data license	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).

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: info@ciesin.columbia.edu

I. Data inputs / methodology

To create this dataset, CIESIN developed a consistent data schema and methodology to harmonize the data from eleven different provinces collected as five different province groups. The province groups are legacy groupings corresponding to the order in which the original data was collected.

- Province group 1: Haut-Katanga, Kasaï, Kasaï-Oriental, Kinshasa, and Lomami
- Province group 2: Haut-Lomami and Tanganyika
- Province group 3: Ituri and Kwilu
- Province group 4: Maniema
- Province group 5: Kasaï-Central
- Province group 6: Tshopo and Mongala

Province Group 1: Haut-Katanga, Kasaï, Kasaï-Oriental, Kinshasa, and Lomami

From January to July 2021 with the support of provincial and national health authorities, local healthcare workers (“head nurses”, “health zone management staff”, and “head doctors of the health zones”) and GRID3 GIS specialists (“mappers” and “provincial coordinators”) engaged in a participatory mapping process in Haut-Katanga, Kasaï, Kasaï-Oriental, Kinshasa, and Lomami. This mapping process occurred at the level of the health zone (an operational unit made up of approximately 15-20 health areas).

Mappers were deployed to health zones in teams of two for approximately nine days where they trained the health area head nurses on data collection using the Geospatial Tracking System (GTS), an Open Data Kit (ODK)-based application. The head nurses routinely work in their respective areas and have a good understanding of the names and location of health facilities, settlements, and points of interest (POIs, such as schools and religious centers) within their health (or catchment) areas.

While the head nurses collected data, the GRID3 mappers worked with the health zone management team to validate and modify data in the field. The GRID3 mappers engaged in a participatory mapping process which included reviewing satellite images with the health zone management team to ensure that the settlement points collected in the field were properly identified, spelled, and in the correct location. They also verified that no settlements were missing. The mappers worked to clean and refine the settlement layer and then presented all of the layers to the health zone head doctor.

Mappers worked with provincial coordinators to consolidate all of the health zones in a provincial settlement layer. These data were sent back to GIS specialists at CIESIN for review province by province between March 2021 and July 2021. GIS specialists identified health areas and health zones with missing

settlement information, removed duplicate points, and ensured the attribute table was correct and consistent.

From July 2021 through October 2022, the mappers and provincial coordinators worked with CIESIN staff to consolidate the data (spelling, gaps and overlaps, topology, etc). This data was used to produce health area level basemaps that were shared back with each health zone and province authorities for a second round of validation.

Mappers working from the GRID3 office in Kinshasa filled in missing information and corrected any identified errors by communicating directly with health zone management staff over the phone and checking original field data. The mappers also compared existing data to the GRID3 settlement extents and in cases where large settlement extents had no associated spatial point, they coordinated with the health zones to find a name for the settlement. Duplicate settlement information and settlement points in areas where too many points were collected for the same settlement (usually in urban areas) were removed. Finally, the mappers compared the settlement points to the microplans to try to identify and locate any missing points.

From November 2022 through January 2023, the in-country GIS team worked with CIESIN staff to integrate these corrections into a final geodatabase.

This work was also done with the participation and supervision of the Direction du Système National d'Information Sanitaire (DSNIS). The Agence Nationale d'Ingénierie Clinique, de l'Information et de l'Informatique de Santé (ANICiiS) also played an important role in the area of data governance. This work was part of the GRID3 Mapping for Health project.

Province Group 2: Haut-Lomami and Tanganyika

Haut-Lomami and Tanganyika settlement data was originally created through an extensive fieldwork exercise spanning from July to September 2019 and supplemented with additional data from the National Malaria Elimination Programme in the DRC (PNLP).

Table 1: Data sources

Source	Description
PNLP	Programme National de Lutte contre le Paludisme (National Malaria Elimination Programme) of the DRC
GRID3	DRC's Ministry of Public Health (Ministère de Santé Publique) and the Center for International Earth Science Information Systems)

Phase 1: 2019 Field data collection

With the support of provincial and national health authorities, local healthcare workers and GRID3 GIS specialists engaged in a participatory mapping process in Haut-Lomami and Tanganyika from July to September 2019. This mapping process occurred at the health zone level.

Mappers deployed to the health zones organized participatory mapping meetings with local healthcare workers, and trained head nurses to collect settlement, health facility, and other points of interest in their respective health areas using an ODK-based smartphone application. Mappers then used this information to delineate health area boundaries. This data was then sent back to CIESIN for additional quality checks.

Phase 2: Integration of data from the PNLP

In October 2021, GRID3 received access to a large geospatial point dataset collected during a bednet distribution campaign by IMA World Health and owned by PNLP. This dataset was combined and consolidated with GRID3's previous settlement data from these two provinces. Only data from PNLP and GRID3 were used. In total, the PNLP data added more than 3,500 points.

Province Group 3: Ituri and Kwilu

Ituri and Kwilu settlement data were created using settlement point data from the PLNP and combining it with additional data sources (see table 2).

Between 2021 and 2022, GRID3 received PLNP data covering the provinces of Ituri and Kwilu. This data consisted of household-level GPS points with village, health area, and health zone attributes. It had been collected by IMA World Health, an implementing partner of the PNLP, during province-wide bed net distribution campaigns. In Ituri, over 1,165,000 household points were received from a bed net distribution campaign conducted in June 2021. In Kwilu, over 1,191,000 household points were received from a bed net distribution campaign conducted in July 2022. These household points were processed and mapped against the GRID3 settlement extents datasets as a way of validation. Further, PLNP-derived settlement points were combined with other data sources in order to compile a geodatabase as complete as possible.

Table 2: Data sources

Acronym	Source Name
OMS GPEI	Initiative Mondiale pour l'éradication de la Poliomyélite (GPEI) & Organisation Mondiale de la Santé (OMS)
PNLP	Programme National de Lutte contre le Paludisme (PNLP)

Acronym	Source Name
ESPK ECV	Ecole de Santé Publique de Kinshasa (ESPK) / Enquête de Couverture Vaccinale (ECV)
RGC	Référentiel Géographique Commun
UCLA	Programme de Recherche et de Formation en Santé, University of California, Los Angeles (UCLA)
CIESIN	Center for International Earth Science Information Network (CIESIN)

In October 2022, two separate field missions (one per province) were organized to verify the data. In Kwilu, budget constraints necessitated verification work be carried out at the antenna level with provincial level staff, rather than directly work at the health zone level. In Ituri, GRID3 mappers worked with local health teams, where preliminary layers were validated and edited.

All modifications from both missions were sent back to CIESIN for final verification before publication. CIESIN ensured that the final layers were free of spelling and topological errors.

Province Group 4: Maniema

The Maniema data was created through an extensive fieldwork data collection conducted by the Kinshasa School of Public Health (Ecole de Santé Publique de Kinshasa, ESPK) and supplemented with additional data from the PLNP.

Table 3. Data sources

Name	Data type/ format	Input data year	Source Acronym
Fieldwork data collected by the Kinshasa School of Public Health (ESPK) in collaboration with GRID3 and CIESIN.	Spatial points	2024	ESPK GRID3
Pre-Distribution Registration Survey (PDRS) from the National Malaria Control Programme (PNLP) collected as part of the anti-malaria campaigns in the Democratic Republic of the Congo	Spatial points	2023	PNLP

Phase 1: Integration of data from the PNLP

In October 2021, CIESIN received access to a large settlement point dataset collected during a bednet distribution campaign by IMA World Health. This dataset was explored, cleaned, and matched against pre-existing microplans from Maniema. Unmatched settlement names (i.e. villages in the microplans but not within the PLNP dataset) were marked as priority settlements to collect during fieldwork to obtain their geographic locations.

Phase 2: Field data collection and data processing

ESPK with the support of provincial and national health authorities, local healthcare workers and GRID3 GIS mappers engaged in data collection from October 2023 to January 2024. Teams of enumerators were deployed to collect settlement, health facility, and POI data using an ODK-based tool. GRID3 GIS specialists then used this information to delineate health area and health zone boundaries. This data was then sent back to CIESIN for final processing and quality checks.

Province Group 5: Kasaï-Central

A comprehensive geospatial survey was conducted by ESPK between March and May 2024 in collaboration with GRID3 and partners. Similar as in Maniema, the survey team collected names, geospatial locations, and relevant attribute information of settlements serviced by health authorities of the Kasaï-Central province.

Table 4 - Data inputs

Source Name/ description	Data type/ format	Input data year	Source Acronym
Fieldwork data collected by the Kinshasa School of Public Health (ESPK) in collaboration with GRID3 and CIESIN.	Spatial points	2024	ESPK GRID3

CIESIN proceeded with data processing, integration, and harmonization steps to create a clean settlement names dataset. Data processing steps included standardization of settlement names and types, as well as clustering to detect anomalies in spatial accuracy and/or attribute information. Visual inspection using Maxar's satellite imagery (2022-2024) was conducted when relevant.

A consistent data schema prioritizing information preservation was also developed to include data fields as per common use cases. A standardized data schema enables seamless integration across subsequent versions of GRID3 settlement names datasets in the DRC.

Province group 6: Tshopo and Mongala

A comprehensive geospatial survey was conducted by ESPK between March and May 2024 in collaboration with GRID3 and partners. Similar as in Maniema, the survey team collected names, geospatial locations, and relevant attribute information of settlements serviced by health authorities of the Tshopo and Mongala provinces

Phase 1: Integration of data from the PNLP

CIESIN extracted and cleaned settlement locations and names contained in the Pre-Distribution Registration Survey (PDRS) data from the National Malaria Control Programme (PNLP), collected in Tshopo and Mongala provinces in 2021 and 2023 respectively. We utilized this preliminary dataset as the baseline during fieldwork.

Table 5 - Data inputs

Source Name/ description	**Data type/ format**	**Input data year**	**Source Acronym**
Pre-Distribution Registration Survey (PDRS) from the National Malaria Control Programme (PNLP) collected as part of the anti-malaria campaigns in the Democratic Republic of the Congo	Spatial points	2021-2023	PLNP
Data collected in the field between March - May 2024 by the Kinshasa School of Public Health in collaboration with GRID3 and CIESIN.	Spatial points	2024	ESPK_GRID3

Phase 2: Field data collection and data processing

A comprehensive geospatial survey was conducted by ESPK between March and May 2024 in collaboration with GRID3 and partners. The survey team collected names, locations and other attribute information about settlements serviced by health authorities in both Tshopo and Mongala provinces.

CIESIN conducted data processing, integration, and harmonization to create an enhanced settlement names dataset. Data processing steps included standardization of settlement names and types, as well as clustering to detect anomalies in spatial accuracy and/or attribute information. Visual inspection using satellite imagery was conducted as needed.

A consistent data schema prioritizing information preservation was also developed to include data fields as per common use cases. A standardized data schema will enable seamless integration across subsequent versions of GRID3 settlement names datasets in the DRC.

II. Dataset Description

The GRID3 COD - Settlement Names v3.0 dataset consists of spatial point data with attribute information (see table 6). The data are available for download in OGC Geopackage format contained in a zip file; metadata files are also included.

Table 6: Codebook

Variable Names	Type	Definition
OBJECTID	numeric	Software- generated unique code
pays	text	Country name French official UN member state spelling
iso3	text	ISO3 code
province	text	DHIS2 province name
prov_uid	text	The identifier of the province in the DHIS2 pyramid, used by the DSNIS
antenne	text	Antenna name
zonestante	text	DHIS2 health zone name
zs_uid	text	The identifier of the health zone in the DHIS2 pyramid, used by the DSNIS
airesante	text	DHIS2 health area name
as_uid	text	The identifier of the health area in the DHIS2 pyramid, used by the DSNIS
localite	text	Settlement name
localitetype	text	Type of administrative entity (locality, avenue, block, etc.)
enclav	text	Site isolation at certain times of the year
enclavdate	text	If isolated, during which quarter of the year?
source	text	Institution or project providing point data for this dataset
date	text	Year of data collection or last edit/modification
lat	numeric	Latitude in decimal degrees
lon	numeric	Longitude in decimal degrees
precision	numeric	Precision of GPS coordinates indicating the tolerance in meters.
grid3id	text	Internal GRID3 ID

III. Known Data Limitations and Disclaimer

The spatial accuracy of the settlement 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 collected data throughout the validation process. In general, it was assumed that the field-collected data were more accurate than secondary sources. Temporal mismatches may exist among this dataset and the satellite imagery used to perform quality checks. This may lead to settlements not being identified or the inclusion of abandoned settlements. Likewise, spelling mistakes and/or mismatches may have occurred due to colloquial variations on how data points are referred to in the field.

This dataset should be considered operational; it has not been fully validated by government officials or ministries.

CIESIN, Columbia University, and its co-authors 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 CIESIN, info@ciesin.columbia.edu.

CIESIN, Columbia University, its co-authors, 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.

IV. Acknowledgments

CIESIN and its co-authors thank the following institutions that provided input data and/or assistance with data production:

Acasus, Switzerland
Agence Nationale d'Ingénierie Clinique et du Numérique de la Santé (ANICNS) , DRC
Bill & Melinda Gates Foundation, USA
Bluesquare, Belgium
Bureau Central du Recensement (BCR), DRC
Caritas, USA
Centers for Disease Control and Prevention (CDC), USA
Direction d'Etudes et Planification (DEP), DRC
Direction des Soins de Santé Primaires (DSSP), DRC
Division du Système National d'Informations Sanitaires (DSNIS), DRC
Division Provinciale de la Santé (DPS) de Kinshasa, Kwilu, Kasaï, Kasaï-Central, Kasaï-Oriental, Lomami, Haut-Lomami, Tanganyika, Haut-Katanga, Ituri, Maniema, Tshopo et Mongala.
Ecole de Santé Publique de Kinshasa (ESPK), DRC
Gavi, the Vaccine Alliance, Switzerland
Geospatial Evaluation and Observation Lab (geoLab), College of William & Mary, USA
Global Affairs Canada (GAC), Canada
Global Good, USA
IMA World Health, DRC

Initiative Régionale de Documentation et d'Accompagnement Communautaire au Développement (IDRAC Sarl), DRC
International Federation of Red Cross and Red Crescent Societies (IFRC), Switzerland
International Medical Corps (IMC), USA
Médecins Sans Frontières (MSF), Switzerland
Ministère de l'Environnement et Développement Durable (MEDD), DRC
Ministère de la Santé publique, Hygiène et Prévention, DRC
Novel-T, Switzerland
Open Street Map (OSM), DRC
PATH, USA
Programme Elargi de Vaccination (PEV), DRC
Programme National de Lutte contre le Paludisme (PNLP), DRC
Référentiel Géographique Commun (RGC), DRC
Soins de Santé Primaires en Milieu Rural (SANRU), DRC
The International Organization for Migration (IOM), DRC
United Nations Children Fund (UNICEF), USA
United Nations Development Programme (UNDP), USA
United Nations Office for Project Services (UNOPS), Denmark and DRC CO
United Nations Office for the Coordination of Humanitarian Affairs (OCHA), USA
United Nations Organization Stabilization Mission in the Democratic Republic of the Congo (MONUCSO), DRC
University of California, Los Angeles (UCLA) DRC Health Research and Training Program, USA
VillageReach, USA
World Health Organization (WHO), Switzerland (HQ), Brazzaville (AFRO), Kinshasa (CO)
World Resources Institute (WRI), USA

Funding for this work was provided by GRID3 under grants INV-044979 GRID3 - Phase 2 Scaling, and FAE/GRID3/001/2024 Soutien à la mise en oeuvre des activités du Projet de Vaccination des Enfants Zéro Dose du Fonds d'Accélération de l'Équité (FAE) en RDC.