

Data Release Notes

Name of the dataset	GRID3 NGA - Roads v1.0
Name of the file	GRID3_NGA_roads_v1_0.gpkg
Date of data release	October 1, 2025
File format	OGC Geopackage
Dataset version	v1.0
Abstract	This document outlines the methodology and data sources used to produce the <i>GRID3 NGA - Roads v1.0</i> dataset. The spatial dataset consists of road and path locations in Nigeria.
Dataset citation	Center for Integrated Earth System Information (CIESIN), Columbia University. 2025. GRID3 NGA - Roads v1.0. New York: Columbia University. https://doi.org/10.7916/b42r-6w17 . Accessed <DAY MONTH YEAR>.
Terms of use	Users are free to download, store, access, use, copy, adapt, transform, alter, arrange, build upon, distribute and transmit this work and any derivative works. Attribution of the source must be provided, and further distribution of this work or derived work must maintain the same terms of data use and license as set forth in this Terms of Use. 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-ShareAlike 4.0 International, CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0) and specified in legal code (https://creativecommons.org/licenses/by-sa/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. Input datasets

Multiple input data were used to create the *GRID3 NGA - Roads v1.0*, see table 1 below.

Table 1. Data inputs

Name	Data Type/ Format	Release Year	Input Data Year	Source Acronym
NGA GRID3 Settlement Extents v3.1	Settlement polygons	2024	2016 - 2023	n/a
Meta (Open Mapping at Facebook)	Road segments	2020	2020	CIESIN / Meta
Overture Maps Foundation	Road segments	2025	2025 and earlier	Overture
GPS Track Data collected during various field campaigns led by the National Primary Health Care and Development Agency (NPHCDA) in Nigeria.	GPS Tracking points	2021-2024	2021-2024	NPHCDA

II. Methodology

Road Data Integration and Cleaning

The GRID3 NGA roads dataset integrates two primary sources: Overture Roads data (obtained from OpenStreetMap) and road feature extractions from satellite imagery conducted by Meta (formally Facebook) in 2020. The Overture data serve as the foundational layer of the dataset and were incorporated in full, without any modifications to the geometries. We found several cases where Meta's extractions were false positives—likely due to interference from clouds or other natural features (see figure 1). In that sense when both datasets spatially overlapped, Overture was preferred. Meta's data were used to fill gaps, complementing the dataset in areas not covered by Overture.

To incorporate Meta's data into a usable road network, topological corrections were necessary. Road network models require a specific topology that defines how the line segments (roads) connect to each other and to the nodes (intersections or endpoints). First, Meta segments were snapped to the nearest Overture segments within a 1- meter tolerance. Next, Meta segments were snapped to or extended toward other Meta segments within a 3-meter tolerance (see figure 2). To enable node creation, vertices were added to the Meta road segments using the *Integrate* tool. These segments were then split at the nodes if they belonged to the same road class.

Figure 1: Visual of a case with misclassified roads. The image on the left is raw data merged from Overture and Meta; the image on the right is after cleaning was performed.

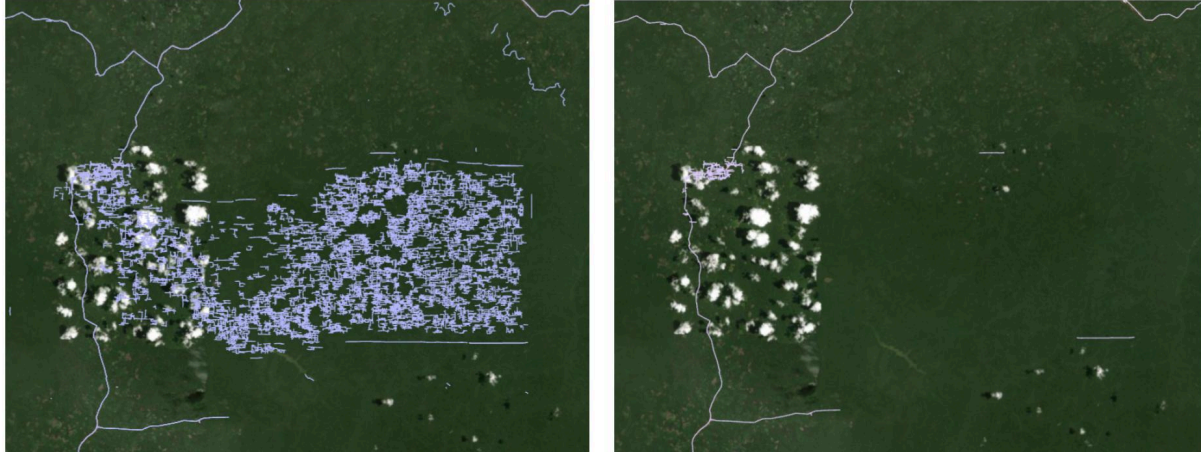
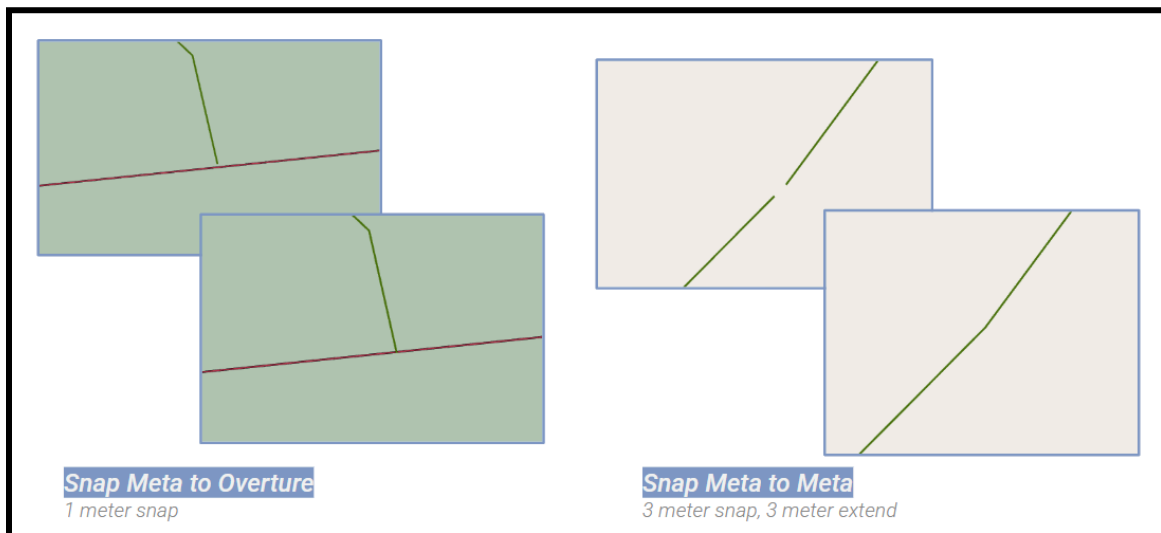


Figure 2: Modifications made to Meta Roads data to correct topology errors..



The Overture and edited Meta roads were merged and buffered by 10 meters to produce continuous line segments. These buffered sections were intersected with GRID3 settlement polygons. Road segments that did not intersect (or touch) a settlement polygon were removed from the final dataset. This step was based on the assumption that roads should connect populated areas, thus unconnected, stand-alone segments may be less likely to represent an actual road. Although the connected road segments still present were valid roads, many were used to get to fields and not for navigation between populated areas. To further improve the data, a network was generated between all settled areas. Only road segments that were part of this network, resided within a settled area, or were a part of the Overture roads were kept. Road segments were also retained if they were necessary to connect the end points of two roads within the cleaned dataset produced in the previous step. To create a more complete network between settled

areas, additional road segments were created to connect settled areas that were disconnected from the road network. These connecting roads were assigned a class of “unknown”. Finally, all roads were clipped to GRID3’s Operational Ward boundaries released in June 2025..

Estimating Travel Speeds from Field-Collected GPS Tracks

Travel speeds represent a practical travel speed for the fastest mode of transportation (walking, motorcycle, or motorized vehicle) for a given road class under low traffic conditions. Travel speeds were assigned by road class, with major road classes calculated using GPS track data, and minor road classes assigned from travel speed literature. GPS track data collected by the National Primary Health Care Development Agency (NPHCDA) between 2021 and 2024 during fieldwork and vaccination campaigns was leveraged to quantify travel speeds for major road classes. Major roads include primary, secondary, tertiary, and trunk. To calculate speed from GPS tracking data, track paths were created between points with a similar device ID that were collected on the same day. Using the time taken to travel between two consecutive points and the distance between points, travel speed was calculated. Track data contained significant noise due to device and human error. Several filters were applied to remove inaccurate driving speeds, including tracks faster than 100 km/h or slower than 2 km/h, distances greater than 10,000 meters, and consecutive ping times longer than 30 minutes.

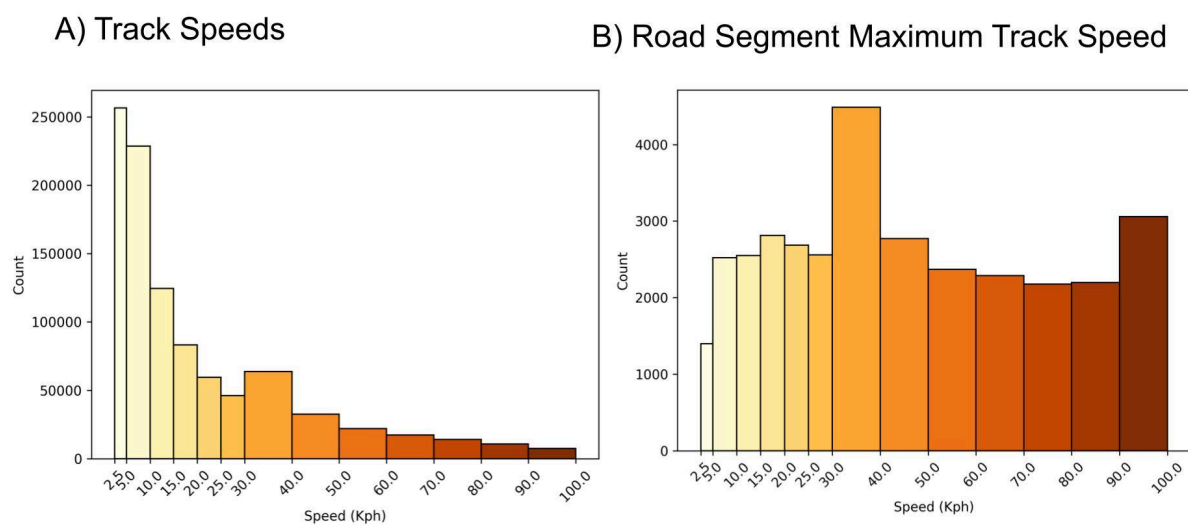


Figure 3: A) Distribution of track speeds for every cleaned and snapped track in Nigeria and B) distribution of the travel speed for cleaned and snapped tracks when spatially joined to road segments using the maximum speed intersecting a road.

Once track paths were cleaned, they were snapped to roads using the *Snap Tracks GeoAnalytics* tool in ArcGIS Pro with a threshold of 250 meters. After cleaning and snapping the tracks, the number of usable points decreased from ~10,000,000 to ~300,000. The cleaned tracks skewed towards slower speeds due to the nature of the track collection campaign (Fig. 3A). To account for slow speeds, track points were spatially joined using a One-to-Many join to the nearest major road segments keeping only the maximum speed of all joined points (Fig. 3B). Generalized speeds for major road classes were determined using descriptive statistics between maximum track speed and road class. Due to the skew towards slower

speeds, the 75th percentile speed value was assigned as the speed for each major road class (Fig. 4). Based on the 75th percentile speed value, primary roads were assigned a speed of 70 km/h, secondary 60 km/h, tertiary 50 km/h, and trunk 80 km/h. These speeds are consistent with speed limits in Nigeria, although generally slower, likely due to traffic and driver/pedestrian behavior (Eneonwo, 2022; OSM, 2025a). Speeds for all other road classes (Table 2) were determined from the literature and OSM's road class descriptions (Keller et al., 2020; Munoz & Källestål, 2012; OSM, 2025b; Ray & Ebener, 2008).

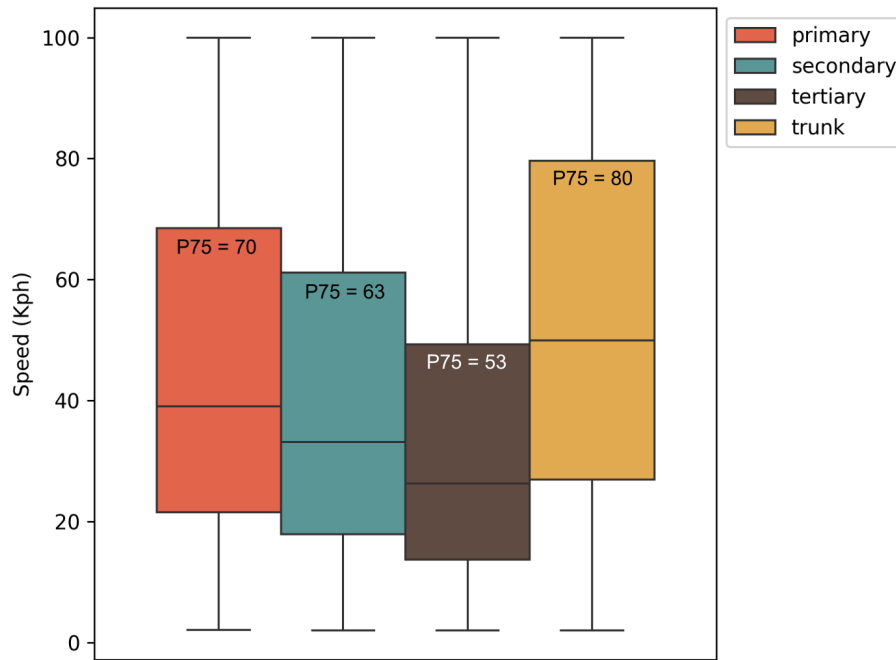


Figure 4: Boxplot of the maximum speeds per road segment grouped by road class with P75 showing the 75th percentile values.

Table 2: Road speeds

Class	Travel Mode	Estimated_speed (km/h)	Estimated_speed_method
motorway	Motorized vehicle	80	GPS
trunk	Motorized vehicle	80	GPS
primary	Motorized vehicle	70	GPS
secondary	Motorized vehicle	60	GPS
tertiary	Motorized vehicle	50	GPS
unclassified	Motorized vehicle	40	literature review
living_street	Motorized vehicle	20	literature review

Class	Travel Mode	Estimated_speed (km/h)	Estimated_speed_method
residential	Motorized vehicle	20	literature review
track	Motorcycle / Walking	10	literature review
footway	Walking	5	literature review
path	Walking	5	literature review
pedestrian	Walking	5	literature review
service	Motorized vehicle	5	literature review
steps	Walking	4	literature review
unknown	Walking	4	literature review

III. Dataset Description

The *GRID3 NGA - Roads v1.0* dataset is a spatial data layer in OGC Geopackage format. The metadata file is included in xml format. The codebook is shown below.

Table 3: Codebook

column_name	column_type	Description
country	text	Country name
iso	text	Country three letter ISO code
id	text	A feature ID. Overture's Global Entity Reference System (GERS) or GRID3 ID assigned to each road segment.
sources_acronym	text	Source of the geometry for the road segment. Acronyms of the organizations, institutions or programs who were involved in the creation of the data input.
class	text	Captures the kind of road and its position in the road network hierarchy.
subclass	text	Specifies the usage of a length of road.
name	text	Names of the road segment.
road_surface	text	Defines the surface material on a road such as paved, asphalt, or unpaved.
speed_limit	numeric	Defines the speed limit of the road segment.

column_name	column_type	Description
speed_estimate	numeric	Estimated travel speed for this segment of road
speed_estimate_method	text	GPS: max speeds from GPS tracks collected during field work : literature review: speed estimates derived from literature review of travel times
source_id	text	Unique identifier referring to each specific data input. The source_id allows users to link each data record with the corresponding input source
date	text	Year of data collection or last edit/modification

IV. Known Data Limitations and Disclaimer

Only minimal topological cleaning was performed during the production of this road dataset. As a result, users should anticipate the presence of gaps between road segments or instances of overlapping geometries. No ground-truthing or independent validation was conducted to verify the accuracy or completeness of the road network.

An additional limitation arises from the process used to intersect roads with GRID3 settlement polygons. When a settlement polygon represents a small, isolated settlement or hamlet, especially one situated far from other populated areas, the likelihood of having a single, continuous road segment extending to it is reduced. Consequently, there is a risk that road segments located en route to such remote settlements were inadvertently removed during the filtering process. Similarly, in cases where uncleaned road segments extended toward a remote settlement, only the segment that directly intersected the settlement polygon was retained, while adjacent but non-touching segments were discarded. This may result in disconnected road fragments appearing across the landscape. Although some of these fragments appear to be entirely unconnected, they were intentionally retained in the dataset at this stage, acknowledging that they may still hold value for future refinements or contextual analysis.

This dataset should be considered operational and not authoritative.

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.

V. References

- Eneonwo, C. (2022). Travel Time Dynamics of the University of Port Harcourt Inter-campus Commercial Transit Services. *International Journal of Advances in Engineering and Management*. DOI: 10.35629/5252-040310971101
- Huerta Munoz, U., & Källestål, C. (2012). Geographical accessibility and spatial coverage modeling of the primary health care network in the Western Province of Rwanda. *International Journal of Health Geographics*, 11(1), 1–11. <https://doi.org/10.1186/1476-072X-11-40/TABLES/2>
- Keller, S., Gabriel, R., & Guth, J. (2020). Machine Learning Framework for the Estimation of Average Speed in Rural Road Networks with OpenStreetMap Data. *ISPRS International Journal of Geo-Information*, 9(11), 638. <https://doi.org/10.3390/IJGI9110638>
- OSM. (2025a). Default speed limits. https://wiki.openstreetmap.org/wiki/Default_speed_limits
- OSM. (2025b). Default speed limits. <https://wiki.openstreetmap.org/wiki/Key:highway>
- Ray, N., & Ebener, S. (2008). AccessMod 3.0: Computing geographic coverage and accessibility to health care services using anisotropic movement of patients. *International Journal of Health Geographics*, 7(1), 1–17. <https://doi.org/10.1186/1476-072X-7-63>

VI. Acknowledgements

Funding for the development and dissemination of this dataset was provided by GRID3 Inc under the Gates Foundation's project INV-044979.