Spatially integrated displacement datasets for Nigeria and DRC, version 1

WorldPop, University of Southampton

Authors: Alexey Noskov, Sarchil Qader, and Attila Lazar. June 2026

Contents

1. Introduction

Forced displacement remains one of the most pressing humanitarian and development challenges in sub-Saharan Africa. Nigeria (NGA) and the Democratic Republic of the Congo (DRC; in the generated datasets we use the standard ISO country code COD) consistently rank among the countries with the highest numbers of internally displaced persons (IDPs), driven by conflict, violence, natural disasters, and other complex crises. Effective humanitarian response and operational planning require not only accurate displacement figures but also geographically disaggregated information describing where people are displaced from and to, the causes of displacement, and how these patterns evolve over time.

This report introduces a spatially explicit, multi-source database of displacement data for Nigeria and the DRC, with the overarching goal of improving the accessibility, interoperability, and analytical utility of subnational displacement estimates. Existing displacement datasets—such as those produced by the International Organisation for Migration's Displacement Tracking Matrix (IOM-DTM), UN OCHA, HDX, and national sources—are often fragmented, heterogeneous, inconsistently structured, and poorly georeferenced. In many cases, displacement records lack robust spatial links to authoritative administrative boundaries, making integration with GIS workflows and other spatial datasets difficult. Differences in naming conventions, administrative coding systems, spatial granularity, and reporting structures further complicate harmonisation across countries and reporting rounds.

To address these challenges, we designed and implemented a reproducible open-source workflow using Tcl scripting, SQLite/SpatiaLite databases, GDAL/OGR libraries, and QGIS. The workflow primarily targets GRID3 administrative boundary layers while maintaining compatibility with HDX-CODAB (Common Operational Datasets – Administrative Boundaries) products. A central objective of the work was to spatially align IDP and returnee records with harmonised administrative units and generate GIS-ready displacement estimates across multiple administrative levels.

The primary operational output of this work is a set of full-wide-format (FWF) GIS layers. In this report, FWF refers to harmonised spatial layers in which multiple reporting rounds, displacement types, origin/destination indicators, and displacement causes are organised in a single wide-format attribute table for each target geography. These layers are intended to be the main user-facing product: they allow regular GIS users to work directly with ready-to-map displacement estimates without querying the full relational database.

The FWF products are built from several underlying spatial linkage products. For Nigeria, the target outputs are ADM2 (LGAs) and ADM3 (Wards) GRID3-based layers. ADM3 destination estimates are primarily based on the *_grd3xynm product, which combines coordinate-based linkage (*_grd3xy) with name-matching linkage (*_grd3). Coordinate-based linkage is preferred where reliable XY site coordinates are available, while name matching is used to improve coverage where coordinates are missing. The suffix xynm therefore indicates a combined XY-and-name-matching product. Nigeria ADM2 layers are then derived by aggregating the corresponding ADM3 outputs. For Nigeria origins, the outputs are based on *_grd3 name-matching products. For the DRC, the main FWF outputs are produced for health zones (ADM3) and health areas (ADM4) using GRID3-based linkage products (*_grd3).

The integration process combines several complementary geocoding approaches. First, where available, administrative identifiers (p-codes) are used to establish direct links between displacement records and administrative boundaries. Second, fuzzy string matching and scoring-based approaches are applied to harmonise inconsistent locality and administrative unit names. To reduce false-positive matches, name matching is geographically constrained using higher-level administrative units whenever possible. Third, GPS coordinates provided in displacement assessments are used to generate coordinate-based spatial estimates, particularly for destination locations. These approaches are combined in a transparent linkage framework that preserves matching metadata and quality indicators for each established spatial relationship.

Displacement datasets are inherently complex and are primarily distributed in tabular form. In this work, these heterogeneous spreadsheets and tables are transformed into integrated relational databases and GIS-ready spatial products. The resulting system provides both detailed relational databases for advanced users and exported GIS layers suitable for operational analysis, mapping, and visualization. The relational database preserves original records, administrative boundaries, spatial links, metadata, and quality diagnostics, while the FWF GIS layers provide simplified, ready-to-use outputs for operational mapping and analysis.

The resulting integrated datasets provide harmonised spatial estimates of IDPs and returnees across multiple administrative levels, including Admin 1–3 units, health zones, health areas, and site-level coordinate products where available. The time coverage spans 2022–2025 and includes both original tabular data and derived GIS layers in multiple export formats. Particular emphasis is placed on preserving transparency, including extensive metadata, matching scores, quality assessment outputs, and linkage diagnostics.

Overall, this work establishes a scalable and reproducible framework for transforming fragmented displacement datasets into harmonised operational geospatial products for Nigeria and the DRC. The integrated database and associated FWF GIS layers are intended to support humanitarian response, operational microplanning, spatial analysis, and future development of displacement estimation methodologies.

This report is organized into nine sections. Section 2 describes the displacement and administrative boundary datasets used, the software tools developed, and the workflow used to assemble the integrated database. Sections 2.1–2.2 introduce the administrative boundary datasets for the Democratic Republic of the Congo and Nigeria, followed by descriptions of the spatial database construction (geo2splite.tcl), displacement data integration (csv2sqlite.tcl), geocoding (geolinks.tcl), and automated export workflows (Sections 2.3–2.8).

Sections 3 and 4 present the resulting spatial datasets for the Democratic Republic of the Congo and Nigeria, respectively, with particular attention to the final FWF GIS layers, origin–destination estimates, displacement reasons, and outputs at multiple administrative levels. Section 5 presents quality assessment procedures and matching statistics for the generated products. Sections 6 and 7 summarize the conclusions and future development directions. Finally, Sections 8 and 9 describe data/software availability and acknowledgements.

2. Data Sources, Software, and Integrated Database

All source datasets included in this work were the most recent versions available as of 29 October 2025. These include administrative boundaries, displacement figures, and auxiliary data downloaded from official sources such as UNHCR, HDX, IOM-DTM, and others. Users should verify with the respective providers if more recent updates have been published since this date. This snapshot ensures consistency across datasets for spatial and temporal analyses presented in the accompanying report.

2.1. Administrative Boundaries for the DRC

For the Democratic Republic of the Congo (DRC), we relied on a combination of official administrative boundary datasets and updated health unit boundaries to enable multi-scale spatial analysis of internal displacement. These boundary layers allowed us to consistently link IDP (Internally Displaced Person) origin and destination records to provinces, territories, and public health service areas across the country. All boundaries were integrated into the unified project database using custom tools developed for this work.

The core administrative boundaries—provinces (Admin 1) and territories (Admin 2)—were obtained from the Humanitarian Data Exchange (HDX, https://data.humdata.org/), where they are published under the "Référentiel Géographique Commun" (RGC) framework maintained by the DRC government and the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). This dataset reflects the official territorial structure of the country, including 26 provinces introduced after the 2015 administrative reform, as well as all subordinate territorial divisions such as villes and rural territories. The files were distributed in standard ESRI Shapefile format and stored under cod/hdx, including associated projection and metadata files. These boundaries were used as the initial spatial reference for processing displacement records, particularly for spatial joins involving aggregated displacement statistics at Admin 1 and Admin 2 levels.

In addition to the administrative hierarchy, we incorporated health unit boundaries to support finer-resolution analysis of IDP dynamics, particularly in eastern DRC where health zones often serve as de facto humanitarian planning units. Health boundary data were sourced from two distinct but complementary providers. The first set was downloaded from HDX's DRC Health Data package (https://data.humdata.org/dataset/drc-health-data), which includes both health zones (zones de santé) and health areas (aires de santé). These boundaries are widely used in public health planning and humanitarian needs assessments and were provided in Shapefile format under cod/hdx/hz.* and cod/hdx/ha.*. These files include attributes such as zone names and identifiers, and were particularly useful for contextualizing displacement patterns within the eastern provinces of Nord-Kivu, Sud-Kivu, Ituri, and Tanganyika.

To improve spatial accuracy and ensure consistency with other GRID3-supported datasets, we also integrated updated health zones and health areas from the GRID3 program. These datasets, retrieved from the GRID3 data portal (https://data.grid3.org), include revised boundaries based on remote sensing, field validation, and alignment with official national health registries. The GRID3 health boundaries cover 15 provinces, including many high-displacement regions. Unlike the HDX shapefiles, these files were provided in GeoPackage format and stored under cod/grd3/hz.gpkg and cod/grd3/ha.gpkg. They were imported into the SQLite database using the same geo2splite.tcl tool used elsewhere in the workflow.

All boundary datasets—HDX RGC for administrative levels and GRID3 for health units—were harmonized within the spatial database by ensuring consistent geometry formats, aligning spatial extents, and unifying attribute schemas (e.g., common IDs and names). This supported consistent querying, visualization, and statistical aggregation. The resulting framework enabled flexible spatial disaggregation, including IDP mapping at the province, territory, health zone, and health area levels. These boundary layers were critical to producing all spatial outputs in this report, including maps of origin and destination patterns, time series charts, and others.

2.2. Administrative Boundaries for Nigeria

To support the spatial analysis of displacement patterns in Nigeria (NGA), we integrated administrative boundary data from two complementary sources: the Office of the Surveyor General of the Federation (OSGOF), accessed via the Humanitarian Data Exchange (HDX), and the GRID3 program. Together, these sources enabled both high-precision geocoding of IDP records and harmonized spatial aggregation across national, state, and local levels.

The OSGOF-derived datasets, retrieved from HDX (https://data.humdata.org/dataset/cod-ab-nga), include official administrative boundaries at levels 0 (national), 1 (states), and 2 (local government areas, or LGAs), with partial coverage of level 3 (wards) in the North-East geopolitical zone. These shapefiles were originally designed for humanitarian use and include both geometry and administrative identifiers (p-codes) aligned with government standards. The OSGOF files were particularly valuable for geocoding, as they exhibited strong compatibility with the IDP data collected through IOM DTM and OCHA, especially in terms of administrative names and p-code references. During the geocoding process, each IDP record was matched to its corresponding administrative unit based on names and p-codes extracted from the OSGOF layers. This approach ensured high matching accuracy and consistency with humanitarian reporting standards.

Following the initial geocoding using OSGOF, the spatial references were translated to GRID3 boundaries. GRID3 provides harmonized and topologically consistent administrative boundaries for all of Nigeria, including full coverage at administrative levels 1, 2, and 3. The GRID3 boundaries were obtained from the GRID3 open geospatial data platform (https://data.grid3.org), and were used as the definitive spatial framework for aggregation, visualization, and modelling throughout this report.

Unlike the OSGOF boundaries, which are distributed as traditional ESRI Shapefiles, the GRID3 datasets are provided in modern geospatial formats: the state (Admin 1) boundaries were received in GeoPackage format. The use of these formats ensured seamless integration into the project's SQLite database, with all boundaries converted and imported using the custom geo2splite.tcl tool.

The decision to use GRID3 boundaries as the target spatial reference system was driven by multiple factors. First, GRID3 offers updated geometries and more consistent topological structures compared to older shapefiles. Second, the boundary layers are aligned with GRID3's population and health datasets, facilitating future work on IDP modelling at high spatial resolution. Third, the inclusion of full Admin 3 coverage across Nigeria enabled analysis at the ward level—a granularity that is not available from OSGOF outside the North-East. Fourth, the GRID3 boundaries are improved by field data collection and harmonised with in-country stakeholders, therefore, considered more accurate and reliable.

To address the challenge of aligning IDP datasets—originally structured according to OSGOF/OCHA administrative boundaries—with the more precise and topologically consistent GRID3 boundary layers, we developed a dedicated algorithm (available at as a part of NGA country driver script, sdbasm_nga.tcl - https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_nga.tcl&ci=tip, see PHASE 35). The need for this tool emerged from significant spatial misalignments and inconsistencies between the two boundary sources. Although the tabular attributes (e.g. admin codes or names) might seem similar or even identical, the underlying geometries often diverge substantially. These discrepancies can result in mismatches during spatial joins or aggregations, introducing significant error in downstream displacement analyses.

The proposed solution uses a GIS-based overlay method to identify the best spatial correspondences between OCHA/OSGOF administrative units and their GRID3 counterparts. It works by importing both boundary sets into a GRASS GIS environment, performing polygon intersections (overlays), and calculating the area of overlap between each pair of features. For each unit in the original dataset (e.g., an OCHA ADM2 polygon), the script selects the GRID3 polygon with which it shares the largest area of intersection. This area-based matching provides a spatially grounded correspondence table that can then be used to transfer displacement data to the new boundary system with minimal loss of fidelity.

Once the best matches are established, the script updates all relevant IDP and returnee tables. New harmonized tables (layers with the '_grd3corr' suffix) are generated by replacing the original administrative IDs with their GRID3 equivalents. This process is applied consistently across all spatial units and displacement tables—whether the data pertains to ADM1/ADM2 levels in Nigeria or Health Zones and Health Areas in DRC. The output is a fully translated database aligned to GRID3 boundaries, enabling more consistent visualization, querying, and spatial aggregation across datasets. This harmonization was critical for ensuring that spatial analyses and derived maps in this report reflect the most accurate and current geographic framework available.

In practice, the workflow for spatial referencing of IDP data followed a two-step approach: initial geocoding using OSGOF shapefiles (due to superior name/p-code match performance), followed by spatial translation to GRID3 boundaries for all subsequent analysis. This approach allowed us to combine the strengths of both sources—leveraging OSGOF's compatibility with input data and GRID3's consistency with planned gridded estimation frameworks.

The integrated boundary layers, now fully aligned with the IDP records, enabled us to produce spatial estimates and maps at Admin 1 (states), Admin 2 (LGAs), and Admin 3 (wards) levels, depending on data availability. This foundation will also support future efforts to disaggregate IDP data to grid cells using covariates such as population density and land use, in alignment with WorldPop methodologies.

In summary, the combined use of OSGOF and GRID3 datasets ensured both accurate geocoding and harmonized spatial analysis. The transition from OSGOF to GRID3 boundaries within the database provides a robust, scalable framework for national and subnational displacement analysis across Nigeria.

2.3. Assembling a Spatial Boundary Database with geo2splite.tcl

To facilitate the integration of administrative and health boundary data into a unified spatial database, we developed a custom lightweight tool called geo2splite.tcl. This script, written in Tcl, automates the ingestion of vector geospatial data—regardless of whether it originates from GeoJSON, ESRI Shapefiles, or GeoPackages—into a SpatiaLite-enabled SQLite database. The tool is openly available at: https://igis.n-kov.com/i.htm/file?name=c/geo2splite.tcl

The goal of this tool is to provide a consistent and minimal interface for loading spatial layers, enriching them with metadata (such as names, identifiers, and additional attributes), and harmonizing geometry types for downstream geospatial analysis and tile-based visualization. The resulting database contains a standardized spatial table named ‘el’ (short for element) that holds all geometries along with key attribute values, zoom levels, and data source information.

The script is invoked from the command line using tclsh and accepts the following inputs:

- A comma-separated list of file paths or glob patterns pointing to input spatial datasets (e.g., *.json, *.shp, *.gpkg).
- The path to the output SQLite database file.
- Optional parameter set by user, a zoom level integer (used to encode spatial resolution).
- Optional parameter, a pair of column names specifying the feature name and its ID (name,id).
- Optional parameter, one or two column names whose values should be retained for reference.

```
USAGE:
 tclsh geo2splite.tcl '/path/to/files/*.json,/path/to/file.json,/path/to/fileN*.shp' result_database.sqlit
e [zoom_level] [name_id_columns] [copy_column_data1,copy_column_data2]

  '/path/to/files/*.json,/path/to/file.json,/path/to/fileN*.shp' - paths to spatial data
                                              separated by comma "," - each one will be
                                              evaluated by the "glob" command (USE single
                                              quotes ' to prevent the subtitution by shell!);

   result_database.sqlite - path to an output database. It will be created if the file does not exist;

   (OPTIONAL) zoom_level - a zoom level integer value (any integer is acceptable, but values from 0-20 are
 recommented for OpenStreetMap-like tile zoom levels, and 0-6 to admin boundaries zoom levels, where 0 is
country level and 5 is settlement level).
   (OPTIONAL) name_id_columns -  comma-separated columns (in shp) or tag names (in json) providing name of
 the feature and it's id. Accetable forms: 'name,id' (name and id columns), 'name' or 'name,' (name column
 only), or ',id' (id column only).).
   (OPTIONAL) copy_column_data1,copy_column_data2 - comma-separated columns to copy data; acceptable forms
: 'col1,col2' 'col1,' 'col1' ',col2'.

 Example (note quotes!):
   tclsh geo2splite.tcl '/home/wdeb/projectsdata/dq/SanDona.OSM.json,/home/wdeb/projectsdata/dq/psi/ontoma
p_renamed/*.json,/home/wdeb/projectsdata/dq/psi/sandonashp_renamed/*.shp /home/wdeb/projectsdata/dq/wgn.sq
lite' 20 name_field,data_id test:data:
   tclsh geo2splite.tcl NGA/src/cod/hdx/nga_adm_osgof_20190417/nga_admbnda_adm1_osgof_20190417.shp test_ng
_db.splite 1  ADM1_EN,ADM1_PCODE NGA/src/cod/hdx/
```

Figure 1 Usage summary for geo2splite.tcl (command line interface)

When executed, geo2splite.tcl performs the following steps:

1. Input Parsing and Initialization
   The script begins by parsing the input arguments. If the output SQLite file does not already exist, it initializes a new database using SpatiaLite and sets up the schema. This includes creating the ‘el’ table, which has fields for an internal ID (newly created), external ID (did; obtained from imported datasets), geometry type, zoom level, name, two optional attribute fields (eone, etwo), and the path to the source dataset (dsrc). A spatial geometry column (geom) is also created using AddGeometryColumn.
2. File Discovery and Preprocessing
   For each file path or glob pattern provided, the script resolves actual files using the glob

command. If a file is not in .geojson or .gpkg format, it is converted on-the-fly to GeoJSON using ogr2ogr, ensuring compatibility with the import routines.

3. Loading GeoJSON Files
For JSON- or GeoJSON-formatted files, the function geojson2sqlite reads the file, parses its features, and extracts geometry, name, ID, and optional attribute values. Each feature is assigned a geometry type code based on its structure (e.g., polygon = 2, multipolygon = 5), and an INSERT SQL statement is constructed. The geometries are inserted using SpatiaLite's GeomFromGeoJSON() function. Features are batch-committed to the database every 1,000 records for performance.

4. Loading GeoPackage Files
If the file is a .gpkg, the script executes a specialized import routine (gpkg2sqlite) that attaches the GeoPackage as a secondary SQLite database, detects the name of the geometry table within it, and transfers records directly into the main database. Geometries are loaded using GeomFromGPB(), and their type is inferred using GeometryType.

5. Geometry Type Mapping and Zoom Levels
Each geometry is assigned a numeric type code corresponding to its type (point, line, polygon, etc.) and tagged with a lod (level of detail or zoom level) value. These fields support downstream filtering and rendering.

6. Cleanup and Finalization
Temporary files created during the preprocessing (e.g., converted GeoJSON files) are deleted after use. The database connection is closed, and the script outputs a status message indicating successful import.

- Database Structure

The final output is a SpatiaLite-enabled SQLite file containing a single core table ‘el’, where each row represents one feature from the input layers. The schema is designed to support both spatial queries and metadata filtering. The consistent use of EPSG:4326 ensures compatibility with most web-mapping and tile-based tools.

- Application in This Project

We used geo2splite.tcl to ingest all administrative and health boundary datasets for both Nigeria and the Democratic Republic of the Congo into a common spatial database. This included Shapefiles from OSGOF and HDX, GeoJSON files from GRID3, and GeoPackages for high-resolution health boundaries. By encoding the level of detail and source path alongside the geometry, the tool enabled robust geospatial joins with IDP data and flexible aggregation at multiple scales (Admin 1–3, health zones, etc.).

The tool’s simplicity and extensibility make it well-suited for rapid updates and future expansion into other countries or displacement contexts. It can also be integrated into automated workflows for large-scale spatial data harmonization, as needed for global or regional displacement monitoring platforms.

2.4. IDP Data for the DRC

To analyse the spatial distribution of internally displaced persons (IDPs) and returnees across the DRC, we compiled and processed displacement datasets published on the Humanitarian Data Exchange (HDX) platform. The core datasets originate from site assessments conducted under the coordination of the United Nations Office for the Coordination of Humanitarian Affairs (OCHA), drawing primarily from data collected by the International Organization for Migration's Displacement Tracking Matrix (IOM DTM) and its partners. These assessments are among the most comprehensive and consistently updated sources of IDP information available for DRC.

All IDP and returnee data were obtained from two main HDX resources:

- IDPs – Site Assessment:

https://data.humdata.org/dataset/drc-displacement-idps-site-assessment-ocha

- Returnees – Site Assessment:

https://data.humdata.org/dataset/drc-displacement-deplace-site-assessment-ocha

The datasets contain structured, population-level information collected through multisectoral assessments conducted in health zones and localities with reported displacement events. The data model employed by OCHA standardizes information on population group type, location (using health zones and administrative boundaries), causes of movement, dates of displacement and assessment, and the status of site occupancy (e.g., host family, displacement site, returnee, etc.). This harmonized model, curated and published by OCHA, builds on raw IOM DTM data but applies cleaning and standardization processes that enable its use in country-wide analyses and spatial aggregations. Unlike direct IOM exports that often vary in format and coverage, the OCHA-processed datasets ensure temporal comparability and full national coverage—critical requirements for spatial modelling of displacement trends.

For this analysis, we reviewed all available post-2022 displacement data to identify a manageable and representative temporal subset. To achieve consistency and avoid seasonal biases, we selected data releases from March and September of each year, starting from March 2022 and extending through March 2025. This semi-annual selection strikes a balance between data volume and temporal sensitivity, while also aligning with operational reporting rhythms. All files were downloaded in XLSX format from HDX and renamed for clarity using the following naming convention:

- idps_YYMM.xlsx – for internally displaced persons
- rets_YYMM.xlsx – for returnees

The corresponding files are stored in the project folder as:

idp/xls/idps_2203.xlsx to idps_2503.xlsx

idp/xls/rets_2203.xlsx to rets_2503.xlsx

Each file was converted to a CSV format for streamlined parsing and integration:

idp/idps_2203.csv to idps_2503.csv

idp/rets_2203.csv to rets_2503.csv

All CSV files share the same column structure (ensured by OCHA), including fields for:

- Movement date, evaluation date, and creation date
- Number of persons and households
- Causes of displacement
- Displacement origin and destination locations (captured at multiple levels: Admin 1–5, health zones, and health areas)
- Hosting arrangements and site locality descriptions

An example of the column structure is:

id,alert_id,movement_date,creation_date,evaluation_date,household,person,...,admin1_id,admin1_label,...,admin1_id_from,admin1_label_from,...

The data structure allows for the reconstruction of population flows at varying administrative and health boundaries. It also facilitates consistent spatial joins with geographies processed through the geo2splite.tcl tool, supporting further disaggregation, mapping, and statistical analysis.

While raw IOM DTM data remains an important source of field-level displacement information, the OCHA-published datasets offer several advantages for a project of this scale:

- Harmonized structure: The data follows a stable schema, making it suitable for systematic multi-temporal analysis.
- Comprehensive coverage: The OCHA-processed records span the entire country, not just IOM's operational areas, by aggregating partner data into a unified format.
- Consistent administrative and health zone matching: The data include standardized p-codes and names that align well with the HDX and GRID3 boundary layers used in this project.
- Validated aggregation levels: The use of health zones and administrative units allows results to be compared across time and provinces.

This approach ensures a robust basis for national-level spatial modelling and comparison, overcoming known gaps and inconsistencies in direct field datasets.

A distinctive feature of this study is the inclusion of returnee data, which has not been systematically integrated into prior spatial displacement studies for DRC and NGA. Returnee records—capturing the number and location of individuals who have returned to their place of origin following displacement—are recorded in the same format and level of detail as IDP data. Their inclusion allows for a more complete picture of displacement dynamics, particularly in conflict-affected eastern provinces where population movement is cyclical and return processes are ongoing.

By integrating both IDP and returnee datasets, this project enables spatial comparisons between ongoing displacement hotspots and areas of reintegration. It also provides the foundation for constructing net displacement balances over time—crucial for informing durable solutions and targeting humanitarian assistance.

The IDP and returnee datasets used in this report provide a reliable, harmonized, and spatially compatible foundation for analysis. Their structure, coverage, and alignment with the project's administrative boundaries ensure that the estimates and visualizations presented throughout this report are based on the most comprehensive displacement data currently available for the DRC. The use of semi-annual data snapshots ensures regular temporal intervals, while the incorporation of returnees expands the scope of analysis to include both active displacement and population recovery patterns.

2.5. IDP Data NGA

For Nigeria, we assembled a comprehensive collection of internally displaced persons (IDPs) and returnee data by leveraging site and baseline assessments from the International Organization for Migration's Displacement Tracking Matrix (IOM DTM). IOM DTM has maintained the most extensive and systematic tracking system for displacement in Nigeria since the onset of the North-East crisis in 2014. The dataset is structured into clearly defined geographical zones and operational rounds, which capture the temporal evolution and spatial patterns of displacement across the country.

All source files were downloaded from the official IOM DTM platform: https://dtm.iom.int/datasets?f%5B0%5D=dataset_country%3A76

We included all available site and baseline assessment rounds from 2022 to 2025 that had sufficient geographic and population detail. The data covers the three main displacement-affected zones in Nigeria:

- North-East (NE): the original epicenter of displacement, including Borno, Adamawa, and Yobe.
- North-Central and North-West (NC/NW): increasingly affected by armed banditry, communal clashes, and flooding.

Each file represents a distinct data collection round or assessment wave. Files are named using a standardized format to indicate zone, data type (IDPs, returnees, or both), assessment type (baseline or site), round number (rN), and the month of publication (YYMM). For example: idps_ncnw_sa_r14_2504.xlsx refers to the Nigeria — North-Central and North-West Zones Site Assessment Round 14, published in April 2025.

After download, each file was converted to CSV format and renamed for compatibility and integration into the spatial database. The final working directory contains 22 CSV files stored under idp/, covering both IDPs and returnees across zones and rounds.

Two principal assessment methodologies are employed in the dataset:

1. Baseline Assessments (BA): These provide exhaustive enumerations of displacement within affected areas, including the number of individuals, households, origin/destination states and LGAs, reasons for displacement, and housing arrangements. Baseline rounds are particularly valuable for capturing static snapshots and cumulative displacement trends.
2. Site Assessments (SA): These focus on living conditions, needs, and vulnerabilities in camps or host communities. While more localized, SA datasets often include richer

metadata (e.g., shelter type, wash availability, education access) and allow spatial linking to smaller administrative units or specific sites.

Both types of assessments are organized into rounds. In the North-East, for example, the dataset includes Site Assessment Rounds 40–49 and Baseline Rounds 41–42. The North-Central and North-West datasets include both Baseline Rounds 8–9 and Site Assessment Rounds 10–16, reflecting more recent engagement in those regions.

The structure of each CSV file follows a consistent schema, typically including fields such as:

- Displacement figures: number of IDPs, households, or returnees
- Location: encoded using administrative names and/or p-codes (Admin 1–3)
- Movement information: date of arrival, origin LGAs, host type
- Reason for displacement and housing type
- Assessment metadata: round number, data collection date

For example, North-East files such as idps_2211r43ne4sa.csv contain detailed displacement observations at both the ward and Local Government Area (LGA) levels, often including information on both origin and current location of IDPs. In contrast, North-Central and North-West files—such as idps_2504r14ncnw4sa.csv—follow a similar structural model but show increased focus on host-community displacement and inter-state movement, sometimes with less granularity in origin data. This variation across regions and assessment rounds illustrates one of the key challenges in building a harmonized database: even when datasets share a general schema, subtle but important differences in content, resolution, and displacement typology require careful alignment.

To address this, we implemented a multi-step harmonization workflow. This included:

Schema mapping to unify column names and data types across files.

Fuzzy string matching to link administrative names to spatial units, particularly when IDs were missing or inconsistent.

Controlled vocabularies to standardize displacement reasons and population group labels.

Relational linking through correspondence tables to match reported units to harmonized COD-AB and GRID3 boundaries.

Automated aggregation scripts to produce consistent spatial summaries at Admin 1–3 levels.

Through this process, all datasets—despite their structural and semantic differences—were integrated into a common spatial-relational format (stored in idps.sdb), enabling uniform querying and visualization across regions and time.

To prepare the data for spatial analysis, all CSV files were normalized and loaded into the SQLite database. Administrative names and p-codes were cross-matched to OSGOF boundaries during geocoding and subsequently reprojected onto GRID3 administrative units for consistency in spatial aggregation.

Unlike the DRC analysis, which relied primarily on OCHA-standardized aggregates of IOM DTM data, for Nigeria we opted to work directly with the original IOM DTM releases. This decision was guided by several factors:

- Finer spatial granularity: The IOM datasets include information down to ward and site level, allowing for analysis at Admin 3 and, where applicable, XY location level.
- Operational zone delineation: IOM provides separate datasets for NE, NC/NW, and GTB zones, which are operationally relevant and support region-specific analysis.
- Consistent attribute structure: The naming and ID conventions in IOM DTM files map well onto OSGOF's official p-code systems, facilitating reliable geocoding.
- Timely updates: IOM DTM is updated more frequently than OCHA aggregates and reflects evolving displacement dynamics, especially in emerging hotspots.

The use of these datasets ensured full geographic coverage of both protracted and recent displacement zones. It also enabled inclusion of previously underrepresented areas in the North-West, such as Zamfara and Sokoto, as IOM has expanded its data collection footprint in recent rounds.

Another key innovation in this analysis is the systematic integration of returnee data from IOM DTM, which has traditionally been underutilized in displacement mapping. Several rounds—especially in 2024 and 2025—explicitly distinguish returnees from currently displaced populations and record their locations, housing situations, and reasons for return.

For example:

- idpsrets_2410r47ne4sa.csv include detailed idp and returnee population counts and destinations.
- The dataset rets_2503r1ncnw4sa.csv marks the first stand-alone returnee assessment for the North-Central and North-West regions.

By incorporating returnee data alongside IDPs, the analysis enables a more complete view of displacement dynamics—capturing not only areas of ongoing crisis, but also zones of potential recovery and return. This dual perspective allows for analysis of net displacement trends and helps identify areas where reintegration may be underway or where secondary displacement is occurring.

The IOM DTM datasets for Nigeria, encompassing both IDP and returnee data across site and baseline assessments, form the core of our displacement analysis. Their detailed structure, zone-specific coverage, and compatibility with Nigerian administrative boundaries make them uniquely well-suited for spatial analysis. The integration of returnee records extends the analytical depth of this study and aligns with broader objectives of capturing the full lifecycle of displacement, from departure through return and reintegration.

2.6. Integrating IDP Data using csv2sqlite.tcl

To facilitate robust, structured storage and high-performance querying of displacement data across multiple rounds and formats, we developed and employed a custom tool named csv2sqlite.tcl. This utility was designed to efficiently ingest large and heterogeneously formatted CSV files — such as those provided by IOM DTM and OCHA — into a unified SQLite database for spatial and statistical analysis. The tool is freely available at: https://igis.n-kov.com/i.htm/file?name=c/csv2sqlite.tcl&ci=tip

The tool was built using Tcl and depends on the sqlite3, csv, and struct::matrix packages. Its main strength lies in its ability to automatically normalize, type-cast, and enrich incoming tabular data while preserving metadata and preparing the content for downstream spatial operations.

While IDP and returnee data for Nigeria and DRC is generally provided in tabular form, the internal structure and formatting of these tables vary substantially across rounds and regions. Field names differ, some values use inconsistent date formats, and column types (e.g., numeric vs. text) are not explicitly specified. Additionally, missing values are coded in non-standard ways (e.g., na, unknown, -), which must be normalized before analysis.

csv2sqlite.tcl addresses these issues by:

- Automatically parsing column names and inferring types (INTEGER, REAL, TEXT, or DATE) based on content.
- Converting inconsistent date formats into standard YYYY-MM-DD values.
- Normalizing missing values to SQL NULL.
- Creating efficient SQLite tables with indexes and accompanying metadata.
- Allowing optional prefixing of table names and control over column renaming.

This approach ensured that CSV files from various sources — including IOM DTM, HDX, and custom processed datasets — could be loaded into a consistent database format and queried uniformly, regardless of their original structure.

```
USAGE:
 tclsh csv2sqlite.tcl '/path/to/file.csv,/path2/to/fileN?.csv,/path/to/files/*.csv' result_database.sqlite
 prefix -n

  '/path/to/file.csv,/path2/to/fileN?.csv,/path/to/files/*.csv' - paths to CSV data
                                               separated by comma "," - each one will be
                                               evaluated by the "glob" command (USE single
                                               quotes ' to prevent the subtitution by shell!);
  result_database.sqlite - resulting database file: if exists, new tables are added to the existing databa
se; if not, new database file is created.

  prefix (OPTIONAL) - add this string prefix to newly created tales; it must be third argument.

   -n (OPTIONAL) - do not alter column names (by defauld, all non-alphanumeric are removed, and the remain
ing string is converted to lowercase); it can be either third or fourth.

 Example (note quotes!):
   tclsh csv2sqlite.tcl '/home/wdeb/init*.csv,/home/wdeb/second/*.csv,/home/wdeb/third.csv' newtables_ out
.sqlite -n
```

Figure 2 Usage summary for csv2sqlite.tcl (command line interface)

The csv2sqlite.tcl tool was used to convert raw CSV files into a structured SQLite database, serving as a critical component of our data integration pipeline. It is invoked via a standard Tcl shell command, specifying a list of input CSV files or patterns, the output database path, and optional flags to control table naming and column normalization. For each CSV file, the tool analyses the header row and infers SQL-compatible data types. It also computes column statistics such as value ranges, missing value counts, null percentages, and text lengths. Various date formats (e.g., ISO, European, or datetime formats) are detected and standardized automatically. Column names are normalized unless explicitly disabled using the -n flag.

Each file is imported as a new table in the SQLite database. The process creates proper SQL structures, inserts each record with null handling and quoting, and builds indexes on all fields to accelerate queries. In parallel, metadata tables (meta_tabs, meta_cols) are populated with schema and summary information. The resulting tables were then joined with spatial geometries using identifiers such as administrative PCodes or health zone IDs. This enabled spatial aggregation at Admin 1–3 levels, health zones, and XY location units.

The design ensured compatibility across years and assessment rounds despite variations in schema. It supported efficient querying, aggregation, and spatial analysis, while preserving full traceability back to the original CSV files. The tool enabled transparent, reproducible processing of large and diverse IDP and returnee datasets for both Nigeria and the DRC.

2.7. Geocoding of IDP Data with geolinks.tcl

As part of the spatial integration workflow, we developed a lightweight but powerful Tcl-based tool named geolinks.tcl, designed to automatically link tabular IDP datasets to geographic units stored in a spatial SQLite database. The tool establishes spatial or textual correspondences between administrative names, IDs, or coordinates from a dataset and corresponding features in a precompiled geometry table (named ‘el’). This is a critical step in ensuring that all displacement records - whether derived from IOM-DTM or OCHA-aggregated assessments - can be georeferenced with high confidence.

The tool is open and available for inspection and reuse at: https://igis.n-kov.com/i.htm/file?name=c/geolinks.tcl&ci=tip

Many displacement datasets for Nigeria and DRC are provided in spreadsheet form, with textual fields indicating administrative units such as state names, province codes, or localities. However, these textual references often vary in spelling, formatting, or completeness, and they frequently require harmonization before reliable spatial joins can be performed.

Furthermore, displacement records may reference either names, PCodes, or latitude/longitude coordinates, depending on the source and year. To accommodate these inconsistencies, geolinks.tcl supports both *attribute-based* and *coordinate-based* geocoding using customizable matching logic and optional mappings.

This functionality allows seamless integration of IDP records with previously prepared boundary geometries in the ‘el’ table, supporting multilevel spatial analysis across Admin 1–3, health zones, or health areas.

geolinks.tcl operates on an SQLite database already populated with geometries using geo2splite.tcl. It adds *correspondence tables* (lnk_*) that link rows in source IDP/returnee datasets to matching features in the el geometry table. Each correspondence record includes a match score and the type of match (name-based, ID-based, coordinate-based, etc.).

The linkage process supports several flexible modes:

- By name: Matching textual administrative names to the name column in the geometry table.
- By ID: Matching code fields (e.g. ADM1_PCODE) to the did column in the geometry table.
- By coordinates: Associating point coordinates with polygon features using spatial containment (Within()).

- With mappings: Allowing user-defined mapping rules to harmonize names or codes across data inconsistencies.

To ensure transparency, we provide extensive metadata alongside the results. For each established linkage in the *lnk_** tables, two additional fields are included: **ctype** and **score**.

- **score** represents the best matching score calculated for a specific linkage.
- **ctype** indicates the type of connection used to establish the match.

The ctype field is stored as an integer representing a binary code composed of the following components:

- **LL –** linkage based on XY site coordinates
- **ID –** linkage based on administrative identifiers
- **Name –** linkage based on administrative unit names
- **UpID –** linkage restricted using an upper-level administrative unit

Currently, the following ctype codes are used:

- **2** (binary: 0b0010, LL,ID*,Name,UpID) → linkage based solely on administrative identifiers.
- **4** (binary: 0b0100, LL,ID,Name*,UpID) → linkage based solely on administrative unit names.
- **12** (binary: 0b1100, LL,ID,Name*,UpID*) → linkage based on names, with the search restricted by an upper-level administrative unit.

The lnk_* tables may also contain negative **ctype** values. These indicate cases where the linkage was established exclusively using site coordinates. The negative sign explicitly marks that no administrative identifiers or names were used in the matching process.

In such cases:

- The absolute value of ctype represents the number of XY points used to derive the final coordinates.
- The score field represents the maximum coordinate range (i.e., the spatial dispersion) among the points used for the linkage.

This convention ensures that coordinate-only linkages are clearly distinguishable from identifier- or name-based matches while preserving information about the underlying spatial reliability.

This tool also supports IDMC-specific records where origins and destinations are encoded as semicolon-separated strings with associated coordinate pairs.

The tool is executed from the command line via tclsh and accepts both mandatory and optional arguments:

```
USAGE:
 tclsh geolinks.tcl dbname.sqlite table_to_link lod dscr out_table_prefix name_col=string|id_col=string et
wo_col=string  [string_map=list]  [eone_col=string] [eone_map=list] [etwo_col=string]  [etwo_map=list]

   dbname.sqlite - target sqlite database.
   table_to_link - table name to be linked to elements in el.
   lod  - level of details in el.
   dscr - data source string compatible with SQLites' LIKE command.
   out_table_prefix - a prefix to add to an output table name.
   name_col=string|id_col=string - set either name or id column name in an input (table_to_link).
   (OPTIONAL) string_map=list - set map list for name_col or id_col (applyed to data in the input table);
              convention (also applied to eone_map and etwo_map): comma separated list, e.g., "NGA,NG,n_g,
NG".
   (OPTIONAL)  eone_col - column name in the input table corresponding to eone in e.
   (OPTIONAL)  eone_map=list - set map list for the corresponding eone data from the input table.
   (OPTIONAL)  etwo_col - column name in the input table corresponding to etwo in e.
   (OPTIONAL)  etwo_map=list - set map list for the corresponding etwo data from the input table.

OR (for the IDMC model - to be refined to make universal)

 tclsh geolinks.tcl dbname.sqlite table_to_link lod dscr out_table_prefix -idmc

Examples:
   tclsh geolinks.tcl  idps_drcnga.sqlite drc_src_idps_idmc_gidd_idd_2312 1  DRC_ idmc_ -idmc

    while executing
"error "${::n-kov::errorHeader} Wrong number of arguments: $argc ($argv). $USAGE""
    invoked from within
"if {$argc < 5} {
    error "${::n-kov::errorHeader} Wrong number of arguments: $argc ($argv). $USAGE"
}"
    (file "geolinks.tcl" line 38)
```

Figure 3 Usage summary for geolinks.tcl (command line interface)

tclsh geolinks.tcl dbname.sqlite input_table lod dsrc out_prefix name_col=colname|id_col=colname [string_map=...] [eone_col=...] [etwo_col=...] ...

- dbname.sqlite: SQLite database containing the el table with geometry data.
- input_table: Name of the data table to be geocoded (e.g. nga_src_idps_iomdtm_r45).
- lod: Level of detail (e.g. 1 for Admin 1, 2 for Admin 2).
- dsrc: Source string identifying the geometry set in the el table (e.g. NGA_ or DRC_).
- out_prefix: Prefix for the output correspondence table (e.g. dest_, orig_, etc.).
- Either name_col=colname or id_col=colname must be specified.
- Optional string_map: A comma-separated list of name corrections (e.g., Zangon Kataf,Zango-Kataf).

To link input data tables (e.g. IDP records) with corresponding administrative boundary geometries, a specialized tool (geolinks.tcl) was used throughout this project. This tool standardizes and spatially aligns entries by matching names or coordinates to boundary layers in a given SQLite database. When linking fields such as the "State" column to Admin 1 geometries, a series of automated steps are performed.

First, the SQLite database is initialized and spatial functions are enabled via the mod_spatialite extension. The tool then determines the matching strategy based on available inputs: if unique identifiers, place names, or geographic coordinates are provided, it sets the appropriate match mode and prepares a scoring system to evaluate the quality of each match.

Name-based matching is carried out through a flexible multi-stage process. It starts with strict matching (exact and case-normalized), then applies a user-defined string correction map (defined in sdbasm_cod.tcl and sdbasm_nga.tcl country drivers), and finally attempts fallback matching using stripped punctuation and diacritics. Each match is assigned a numeric score, which is stored in a correspondence table. When coordinates are available—such as in IDMC or IOM datasets—the tool performs spatial containment checks to link each point to its enclosing administrative unit. This process handles both origin and destination locations.

For every link operation, a correspondence table is generated. This includes references to the matching geometry, the original data row, the match type, and the associated confidence score. At the end of each run, the tool reports the distribution of scores to help assess matching quality.

This approach allowed efficient, scalable geocoding of large datasets across DRC and Nigeria. It was essential for resolving inconsistencies in administrative names between sources and for assigning accurate spatial references to all IDP and returnee records.

The geolinks.tcl tool plays a critical role in transforming raw displacement tables into spatially-referenced datasets. It supports robust, customizable geocoding that adapts to inconsistent, noisy, or fragmented data — a frequent challenge in humanitarian displacement reporting. It also ensures transparency and traceability by preserving match scores and correspondence metadata, providing a rigorous foundation for all subsequent spatial analysis in this study. In the current implementation, we use a matching score threshold of 1 for NGA and 2 for COD, corresponding to maximum Levenshtein distances of 1 and 2, respectively.

2.8. Automated Assembly and Summary of Integrated IDP Data

All obtained data have been automatically integrated into a single-file SQLite database named 'idps.sdb'. To facilitate this process, we developed three open-source tools, which are freely available under the MIT license. These tools are currently hosted at https://igis.n-kov.com/i.htm/dir?ci=tip&name=c. The tools include:

- geo2splite.tcl: This tool is designed for importing COD-AB shapefiles (it supports most OGR formats) into the SQLite database, allowing for easy handling and analysis of administrative boundaries.
- csv2sqlite.tcl: This tool enables the importation of IDP data tables into the database, ensuring that the displacement data is structured and accessible.
- geolinks.tcl: This tool establishes connections between the IDP data and the corresponding COD-AB rows, facilitating the integration of spatial and displacement information.

These tools provide a streamlined, automated process for assembling and linking the various datasets into a cohesive database, enabling efficient analysis of IDP patterns and trends.

The developed tools allowed us to prepare an integrated, holistic SQLite database. It was created using geo2splite.tcl and consists of an 'el' table containing all COD-AB data. The 'idp_*' tables were imported with csv2sqlite.tcl as separate tables. The 'lnk_*_orig_*' and ‘lnk_*_dest_*' tables were produced with geolinks.tcl and linked the rows of the 'el' and 'idp_*' tables. A simplified database schema is presented in Figure 4. These tools are universal and can be applied to other countries.

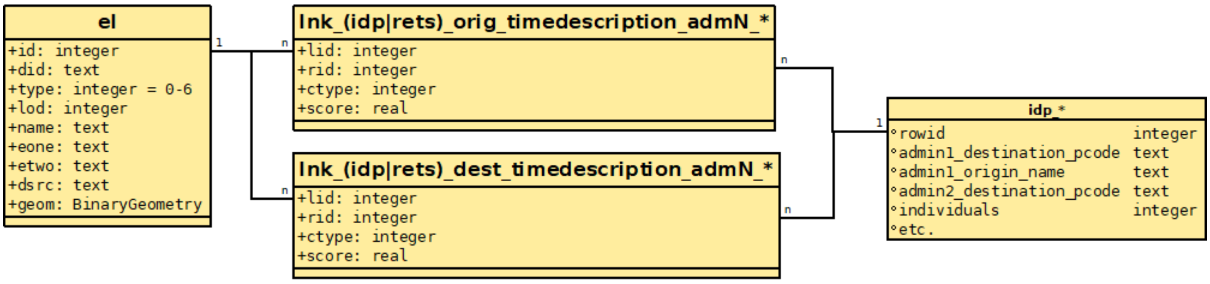

Figure 4. Relational Data Model of the Assembled Holistic Database

To enable unified, multi-source spatiotemporal analysis of IDP and returnee datasets in Nigeria (NGA) and the Democratic Republic of Congo (DRC), we developed a reproducible batch workflow that assembles and links all data into a single spatial database. This integrated database, named idps.sdb, contains harmonized boundary geometries (from GRID3 and HDX), cleaned tabular displacement records (from IOM-DTM and other sources), and all necessary spatial correspondences between them.

The assembly process is fully scripted using a Tcl workflow (see NGA and DRC country drivers available here https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_nga.tcl&ci=tip and here https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_cod.tcl&ci=tip). The workflow scripts employ several modular tools described earlier—geo2splite.tcl, csv2sqlite.tcl, and geolinks.tcl—to construct and link the database in a step-by-step, repeatable fashion. Both reproducible scripts provide detailed documentation and can be used as a basis for other country drivers.

The Nigerian component begins with the integration of subnational boundaries from two major sources:

- GRID3 Nigeria boundaries for Admin 1, 2, and 3 levels (statename, lganame, wardname), and optionally XY locations.
- HDX (OSGOF) boundaries in shapefile format, including detailed PCodes and official names for Admin 1–2 levels.

Each of these geographic layers was processed using the geo2splite.tcl tool to populate the el geometry table in idps.sdb.

Next, all cleaned IDP and returnee datasets from IOM-DTM were imported using the csv2sqlite.tcl tool. This included baseline and site assessments across the North-East, North-Central, and North-West zones, covering multiple rounds from 2022 to 2025. In total, over 20 cleaned and standardized CSV files were converted to structured database tables.

Once tabular and spatial data were present, we applied the geolinks.tcl tool to establish spatial correspondences for each data table. Using PCodes (e.g., StatePcode, LGAPcode) and string-mapped matching (e.g., "NGA" → "NG"), destination locations in each dataset were geocoded to Admin 1 and Admin 2 levels. This approach was repeated for each round and geographic level to ensure consistent spatial referencing across time.

In the case of DRC, a similar approach was used but extended to support health geographies in addition to administrative boundaries. The database includes:

- GRID3 and HDX boundaries for Admin 1 and 2 (provinces and territories).

- Health Zones (HZs) and Health Areas (HAs) integrated at levels 3 and 4 respectively.

These were assembled using geo2splite.tcl, with key fields such as admin1_id, health_zone_id, and their hierarchical relationships.

Displacement records from DTM site assessments and returnee tracking were loaded using csv2sqlite.tcl. Each round—e.g., idp_idps_2503, idp_rets_2411—was linked to corresponding spatial units using geolinks.tcl, handling both origins and destinations at up to four geographic levels.

The country driver script loops over all available DTM rounds to geocode both IDPs and returnees using appropriate ID fields (e.g., admin2_id_from), automating the linkage of over 30 distinct displacement tables.

The final SQLite database, idps.sdb, contains:

- All relevant administrative and health boundaries for DRC and NGA (levels 1–4).
- Cleaned and harmonized tabular data from IOM-DTM and other sources.
- Spatial correspondence tables (lnk_*) linking each row of IDP data to matching polygons in the el table, with match types and scores recorded.
- Metadata tables describing the source files, table structure, and column-level properties.

This integrated database serves as the central data engine for all spatial estimation and spatial outputs and QA described in the remainder of this report. It ensures consistency, reproducibility, and the ability to update with new displacement data rounds rapidly.

The integrated database for the Democratic Republic of the Congo (DRC) contains a wide range of internally displaced person (IDP) and returnee datasets, compiled primarily from IOM-DTM site assessments, aggregated reports, and derived tabular summaries. The datasets span from early 2022 to early 2025, capturing key trends in displacement and return over time.

A total of 22 tables were imported for DRC, comprising both detailed displacement records and aggregated summaries at health zone and provincial levels. Together, they include over 430,000 records and more than 700 MB of CSV source data.

Table 1Summary Table of Integrated DRC IDP Data

Table Name	File Name	File Size (Mb/bytes)	Number of Columns	Number of Rows
idp_idpsagrhz_2409	idp/idpsagrhz_2409.csv	126	34	519
idp_idpsagrprov_2409	idp/idpsagrprov_2409.csv	1	5	27
idp_idps_2203	idp/idps_2203.csv	11448	45	32136
idp_idps_2303	idp/idps_2303.csv	14188	45	35409
idp_idps_2309	idp/idps_2309.csv	22049	46	54865
idp_idps_2403	idp/idps_2403.csv	22418	50	48004
idp_idps_2411	idp/idps_2411.csv	16460	50	36834
idp_idps_2502	idp/idps_2502.csv	16054	40	45280
idp_idps_2503	idp/idps_2503.csv	16551	50	35773
idp_rets_2203	idp/rets_2203.csv	2384	45	6742
idp_rets_2209	idp/rets_2209.csv	3812	45	10183

idp_rets_2303	idp/rets_2303.csv	3123	45	7933
idp_rets_2309	idp/rets_2309.csv	6411	46	17029
idp_rets_2403	idp/rets_2403.csv	4935	50	10735
idp_rets_2411	idp/rets_2411.csv	5164	50	11790
idp_rets_2502	idp/rets_2502.csv	3359	40	9632
idp_rets_2503	idp/rets_2503.csv	3580	50	7855
idp_idps_2504	idp/idps_2504.csv	12	34	33529
idp_rets_2504	idp/rets_2504.csv	2.4	34	7854
Idp_idps_2505	idp/idps_2505.csv	16	42	35774
Idp_idps_2507	idp/idps_2507.csv	7.9	34	26651
Idp_rets_2507	idp/rets_2507.csv	2.6	34	9199

- Temporal coverage: 2022–2025, with approximately two major rounds per year.
- Geographic granularity: Data are structured to support linking at multiple levels: province (Admin 1), territory (Admin 2), health zone (Admin 3), and health area (Admin 4).
- Thematic scope: Includes population counts, causes of displacement, vulnerability indicators, origin and destination attributes, and returnee tracking.
- Aggregated layers: The September 2024 files (idpsagrhz_2409, idpsagrprov_2409) provide health zone/province summaries used for comparison and high-level analysis.

This structured integration supports consistent geospatial analysis of displacement and returnee dynamics over time, including the evaluation of trends, hotspot detection, and administrative-level profiling. The idp_idps_2503 table is a detailed and high-volume displacement dataset compiled for the Democratic Republic of the Congo (DRC), reflecting information as of March 2025. It includes 35,773 rows and 50 columns, with full coverage (0% nulls) across all fields. The table contains structured data on internally displaced persons (IDPs), disaggregated by origin and destination geographies, population counts, displacement causes, and metadata about the reporting sources. It has been fully integrated into the SQLite database and geocoded to administrative units.

The dataset is structurally complete, although several fine-resolution geographic fields are sparsely populated depending on source availability. Each record is uniquely identified by the id field, with an additional id_detail field used for internal tracking. Movement records are grouped using alert_id. The key date fields—movement_date, creation_date, and evaluation_date—are stored as Unix timestamps and are fully populated and consistent. Population information is provided through household and person columns, with household counts ranging up to 24,048 and individual counts up to 168,348. The displaced population is further classified using the population_id and population_label fields (e.g., IDPs, returnees).

Health service geography is encoded via health_zone_id and health_area_id (10- and 14-character codes), with their descriptive counterparts in health_zone_label and health_area_label (up to 22 characters). The type of accommodation or hosting is described using hebergement_id and hebergement_label, and the reason for displacement is recorded in cause_id and cause_label, with detailed causes such as "armed conflict" or "flooding"—some labels extending up to 38 characters.

Geographical disaggregation spans administrative levels 1 to 5 for both origin and destination. All _id fields conform to national PCode standards (4 to 14 characters), while the _label fields provide readable location names (2 to 30 characters). Coverage at Admin 5 level is limited, present in about 3% of records.

Origin information is further enriched with hz_id_from and ha_id_from (original health zones and areas), along with descriptive fields hz_label_from and ha_label_from. Site-level context is provided through description (free text, up to 1,900 characters) and site_localite (locality name, 2–61 characters). Finally, metadata includes primary_source and secondary_source, both consistently populated and reaching up to 234 characters in length. The acronym_fr field captures the reporting actor (e.g., IOM), while actor_Id provides a numeric reference to the contributing organization

Table 2 idp_idps_2503 column overview (March 2025, DRC)

Column Name	Data Type	Precentage of Null Values	Length (in Characters)	min value or length	max value or length
id	INTEGER	0	214632	336002	815412
alert_id	INTEGER	0	1212	4129	5794
movement_date	DATE	0	572346	1.55E+09	1.55E+09
creation_date	DATE	0	572346	1.63E+09	1.63E+09
evaluation_date	DATE	0	572346	1.7E+09	1.7E+09
household	INTEGER	0	56046	1	24048
person	INTEGER	0	80416	1	168348
population_id	INTEGER	0	35772	1	2
population_label	TEXT	0	255349	7	8
health_zone_id	TEXT	0	357720	10	10
health_area_id	TEXT	0	470960	14	14
health_area_label	TEXT	0	232834	3	22
health_zone_label	TEXT	0	232733	3	17
hebergement_id	INTEGER	0	35772	2	5
hebergement_label	TEXT	0	595646	4	17
cause_id	INTEGER	0	35772	1	7
cause_label	TEXT	0	1082190	5	38
admin1_id	TEXT	0	143088	4	4
admin1_label	TEXT	0	265720	5	14
admin2_id	TEXT	0	214632	6	6
admin2_label	TEXT	0	205905	3	15
admin2_type	INTEGER	0	35772	3	4
admin3_id	TEXT	0	254856	8	8
admin3_label	TEXT	0	275616	2	30
admin4_id	TEXT	0	304190	10	10
admin4_label	TEXT	0	259523	3	23
admin5_id	TEXT	0	1050	14	14
admin5_label	TEXT	0	611	4	20
admin1_id_from	TEXT	0	143088	4	4
admin1_label_from	TEXT	0	262946	5	14

admin2_id_from	TEXT	0	214416	6	6
admin2_label_from	TEXT	0	210178	3	15
admin2_type_from	INTEGER	0	35736	3	4
admin3_id_from	TEXT	0	242304	8	8
admin3_label_from	TEXT	0	285893	2	30
admin4_id_from	TEXT	0	291900	10	10
admin4_label_from	TEXT	0	257607	3	23
admin5_id_from	TEXT	0	1008	14	14
admin5_label_from	TEXT	0	645	4	17
hz_id_from	TEXT	0	339210	10	10
hz_label_from	TEXT	0	224168	3	14
ha_id_from	TEXT	0	11438	14	14
ha_label_from	TEXT	0	5746	4	14
description	TEXT	0	951451	3	1900
site_localite	TEXT	0	272431	2	61
id_detail	INTEGER	0	214632	337351	818851
primary_source	TEXT	0	1678808	6	76
acronym_fr	TEXT	0	108041	2	18
actor_Id	INTEGER	0	107291	1	100542
secondary_source	TEXT	0	109667	2	234

The dataset is designed for granular spatial analysis, with administrative coverage extending to Admin 4 or 5 where data permit. The high fill rate and consistency across fields support reliable downstream processing, including linkage to geospatial boundaries and temporal comparison.actor_Id and source fields provide traceability, supporting transparency and methodological reproducibility.

The Nigeria dataset contains internally displaced persons (IDPs) and returnee records collected from IOM-DTM assessments covering the North-East (NE), North-Central and North-West (NCNW) zones, and the Gombe-Taraba-Bauchi (GTB) region. Data come from both Baseline Assessments (BA) and Site Assessments (SA), across 2022–2025. The processed database includes 22 distinct CSV tables, converted to SQLite, and integrated with administrative boundaries for spatial analysis at multiple levels.

Table 3 Summary Table of Integrated NGA IDP Data

Table Name	**File Name**	**File Size (Mb)**	**Columns**	**Rows**
idpsrets_dest_2410r47ne4sa	idp/idpsrets_dest_2410r47ne4sa.csv	11464	510	2996
idpsrets_dest_2503r48ne4sa	idp/idpsrets_dest_2503r48ne4sa.csv	1130	63	2961
idpsrets_dest_2503r49ne4sa	idp/idpsrets_dest_2503r49ne4sa.csv	10169	544	2952
idps_dest_2201r8ncnw4ba	idp/idps_dest_2201r8ncnw4ba.csv	990	87	1564
idps_dest_2206r9ncnw4ba	idp/idps_dest_2206r9ncnw4ba.csv	254	59	860
idps_dest_2307r10ncnw4sa	idp/idps_dest_2307r10ncnw4sa.csv	455	44	1691

Table Name	File Name	File Size (Mb)	Columns	Rows
idps_dest_2307r11ncnw4sa	idp/idps_dest_2307r11ncnw4sa.csv	4105	325	1759
idps_dest_2405r12ncnw4sa	idp/idps_dest_2405r12ncnw4sa.csv	1689	109	1638
idps_dest_2407r13ncnw4sa	idp/idps_dest_2407r13ncnw4sa.csv	622	47	1647
idps_dest_2504r14ncnw4sa	idp/idps_dest_2504r14ncnw4sa.csv	630	50	1734
idps_dest_2503r16ncnw4sa	idp/idps_dest_2503r16ncnw4sa.csv	388	37	1762
idps_dest_2207r41ne4ba	idp/idps_dest_2207r41ne4ba.csv	477	21	2521
idps_dest_2207r42ne4ba	idp/idps_dest_2207r42ne4ba.csv	591	21	2521
idps_dest_2207r40ne4sa	idp/idps_dest_2207r40ne4sa.csv	583	235	300
idps_dest_2211r43ne4sa	idp/idps_dest_2211r43ne4sa.csv	568	35	2451
idps_dest_2305r44ne4sa	idp/idps_dest_2305r44ne4sa.csv	2793	127	2483
idps_dest_2404r45ne4sa	idp/idps_dest_2404r45ne4sa.csv	527	37	2349
idps_dest_2408r46ne4sa	idp/idps_dest_2408r46ne4sa.csv	661	37	2334
rets_dest_2503rXgtb4sa	idp/rets_dest_2503rXgtb4sa.csv	52	78	150
rets_dest_2503r1ncnw4sa	idp/rets_dest_2503r1ncnw4sa.csv	259	79	669
rets_dest_2207r41ne4sa	idp/rets_dest_2207r41ne4sa.csv	217	70	689
rets_dest_2211r43ne4sa	idp/rets_dest_2211r43ne4sa.csv	156	35	706

Temporal coverage: The NGA data spans from early 2022 (Round 8) to March/April 2025 (Rounds 48–49), offering consistent temporal resolution across most quarters. Zonal representation: All major displacement regions—NE, NCNW, and GTB—are represented. GTB, as a focus region, has limited returnee data. Assessment types: Both Baseline Assessments (used early on or in new areas) and Site Assessments (more granular tracking) are included. Returnee tracking: Several rounds now integrate both IDPs and returnees (e.g., Rounds 47–49 for NE), reflecting expanded scope. Column variability: File structure varies significantly between rounds: Early BA rounds: ~20–80 columns. Recent SA rounds: up to 544 columns (e.g., idpsrets_dest_2503r49ne4sa), including full origin-destination details, site info, and source metadata. Record volume: Most rounds contain between 1,500 and 2,500 rows. Combined, this produces a comprehensive picture of displacement dynamics in NGA.

The idp_idpsrets_ne_sa_r48_2503 table includes 2,961 records and 63 columns from IOM-DTM's Round 48 Site Assessment in North-East Nigeria (March 2025). It contains geocoded data on IDPs and returnees at the site level, with full administrative referencing via StatePcode, LGAPcode, and WardPcode.

Key population metrics include NumberofHouseholds, NumberofIndividuals, and year-specific arrival fields for 2021–2024, enabling temporal analysis. Displacement causes and return reasons are recorded in fields such as col15g2ReasonofdisplacementofmajorityofIDPsReturneesinthesite and col15g3ReasonforReturning.

Demographics are well covered with age–sex disaggregation (MaleInfants1y, FemaleAdults1859y, etc.), and summary counts in TotalMale and TotalFemale. Most entries are complete, though origin information (InitiallydisplacedState, InitiallydisplacedLGA) is sparsely filled (around 24%).

Overall, this table supports disaggregated spatial and temporal analysis of displacement and returnee patterns in Nigeria, with a focus on population structure and site characteristics.

Table 4 column overview of idp_idpsrets_ne_sa_r48_2503

Column Name	**Data Type**	**Percentage of Null Values**	**Length (in Characters)**	**min value or length**	**max value or length**
DTMround	TEXT	0	8880	3	3
PopulationGroup	TEXT	0	75495	9	33
HowwasthisLocationassessed	TEXT	0	36303	12	21
Isthelocationfullyaccessible	TEXT	4	46012	16	20
site_accessibility	TEXT	22	6843	2	3
State	TEXT	0	16527	4	7
StatePcode	TEXT	0	14800	5	5
LGA	TEXT	0	18888	3	14
LGAPcode	TEXT	0	23680	8	8
Ward	TEXT	0	28923	3	36
WardPcode	TEXT	0	32560	11	11
SiteType	TEXT	0	42659	4	28
SiteClassification	TEXT	0	41031	7	16
SiteIDSSID	TEXT	0	22164	3	27
SiteName	TEXT	0	29233	3	40
SiteStatusIfCamp	TEXT	95	1224	6	14
SiteStartDate	TEXT	5	28170	10	10
col11f5SiteClass	TEXT	1	14610	5	5
col15g2Reasonofdisplacement ofmajorityofIDPsReturneesinthesite	TEXT	0	31313	10	25
col15g3ReasonforReturning	TEXT	77	12801	14	47
col13g4Settlementsitetype	TEXT	0	71565	12	41
NumberofHouseholds	INTEGER	0	7034	1	33797
NumberofIndividuals	INTEGER	0	9311	3	106339
IDPSRETURNEESHHsArrivalin2024	INTEGER	65	1340	1	2182
IDPSRETURNEESINDsArrivalin2024	INTEGER	65	2048	2	4131
IDPSRETURNEESHHsArrivalin2023	INTEGER	65	1452	1	7072
IDPSRETURNEESINDsArrivalin2023	INTEGER	65	2165	2	22151
IDPSRETURNEESHHsArrivalin2022	INTEGER	57	1986	1	4016
IDPSRETURNEESINDsArrivalin2022	INTEGER	57	2877	2	14731

IDPSRETURNEESHHsArrivalin2021	INTEGER	48	2481	1	9085
IDPSRETURNEESINDsArrivalin2021	INTEGER	48	3624	2	29588
IDPSRETURNEESHHsArrivalBefore 202120142020	INTEGER	4	6496	1	18169
IDPSRETURNEESINDsArrivalBefore 202120142020	INTEGER	4	8639	3	59175
ReturneeStatus	TEXT	0	18257	4	16
initiallydisplacedCountry	TEXT	76	4886	4	8
InitiallydisplacedState	TEXT	76	4129	4	8
InitiallydisplacedLGA	TEXT	76	4726	3	14
InitiallydisplacedWARD	TEXT	76	6290	3	34
MajorityYearofReturns2024HH	INTEGER	77	785	0	404
MajorityYearofReturns2024IND	INTEGER	77	1029	0	3640
MajorityYearofReturns2023HH	INTEGER	76	835	0	183
MajorityYearofReturns2023IND	INTEGER	76	1058	0	1468
MajorityYearofReturns2022HH	INTEGER	76	885	0	3861
MajorityYearofReturns2022IND	INTEGER	76	1093	0	14731
MajorityYearofReturns2021HH	INTEGER	76	965	0	2082
MajorityYearofReturns2021IND	INTEGER	76	1218	0	13533
MajorityYearofReturns20142020HH	INTEGER	76	1979	0	6132
MajorityYearofReturns20142020IND	INTEGER	76	2537	0	49056
MaleInfants1y	INTEGER	0	4693	0	2355
FemaleInfants1y	INTEGER	0	4912	0	2807
MaleChildren15y	INTEGER	0	5759	0	6312
FemaleChildren15y	INTEGER	0	6000	0	7465
MaleChildren612y	INTEGER	0	6112	0	11439
FemaleChildren612y	INTEGER	0	6304	0	13456
MaleYouth1317y	INTEGER	0	6246	0	13414
FemaleYouth1317y	INTEGER	0	6441	0	15754
MaleAdults1859y	INTEGER	0	6961	0	13956
FemaleAdults1859y	INTEGER	0	7109	0	17327
MaleElderly60y	INTEGER	0	4420	0	2978
FemaleElderly60y	INTEGER	0	4491	0	1314
TotalMale	INTEGER	0	8248	1	48377
TotalFemale	INTEGER	0	8523	1	57962

3. Result Datasets for DRC

To generate spatial outputs for the Democratic Republic of the Congo (DRC), a custom Tcl workflow script (https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_cod.tcl&ci=tip) was developed to transform the integrated relational database into GIS-ready spatial products. The workflow uses the harmonised SQLite/SpatiaLite database (idps.sdb) as the primary source and automatically generates exported GIS layers in multiple formats, including GeoJSON, GeoPackage, CSV, and SpatiaLite/SQLite.

The script initializes a new SpatiaLite database (idps_export.sqlite), enables spatial functions through the mod_spatialite extension, and attaches the source database for processing. It then identifies all linkage tables with names beginning with lnk_, representing established spatial correspondences between IDP/returnee records and geographic entities such as provinces, territories, health zones, and health areas.

For each reporting round and linkage table, the workflow automatically generates aggregated spatial layers containing:

• Element identifiers (elid, eldid, elname)

• Total numbers of displaced persons or returnees (num)

• Cause-specific displacement counts (c1–c7)

• Polygon geometries stored in EPSG:4326 (WGS 84)

The aggregation process uses SQL joins between linkage tables (lnk_*), original displacement tables (idp_*), and harmonised administrative boundaries stored in the el table. This enables automated calculation of displacement totals and reason-specific summaries for each geographic unit.

The resulting datasets are exported into several output directories:

• json_export/ – GeoJSON files

• gpkg_export/ – GeoPackage files

• csv_export/ – CSV exports

• fwf_export/ – Full Wide-Format (FWF) products

• idps_export.sqlite – combined SpatiaLite database containing all generated GIS layers

In total, 284 GIS layers were generated for DRC, representing internally displaced persons (IDPs) and returnee flows across multiple reporting rounds and administrative levels. These outputs include both origins and destinations of displacement events and support analysis at Admin 1–2 levels, Health Zones (Admin 3), and Health Areas (Admin 4).

Each exported layer follows a standardized naming convention:

<type>_<direction>_<time>_<admin>_<source>

where:

• type = idp or ret

• direction = orig or dest

• time = reporting period or assessment round

• admin = administrative level (adm1, adm2, adm3, adm4)

• source = spatial linkage approach (hdx, grd3, grd3corr, grd3xy, grd3xynm, or xy)

This convention ensures systematic organization of all displacement layers and compatibility with the integrated GRID3 spatial workflow.

In addition to per-round GIS layers, the workflow also generates Full Wide-Format (FWF) products located in the fwf_export/ directory. These datasets represent the primary operational outputs of this work. Unlike individual assessment layers, the FWF products combine all reporting rounds into compact GIS layers where displacement estimates from different time periods are stored as columns within a single dataset. This structure simplifies temporal comparison, operational microplanning, and external data integration.

For DRC, the FWF outputs include:

• adm3.* – Health Zone–level datasets

• adm4.* – Health Area–level datasets

The FWF products are exported in CSV, GeoJSON, GeoPackage, and SQLite/SpatiaLite formats. Additional *.nogeo.csv tables are provided without geometries to facilitate use in spreadsheet software such as Microsoft Excel or LibreOffice Calc.

The generated spatial products can be readily visualized and analysed using standard GIS software such as QGIS. Several demonstration maps and layouts were prepared as part of this project to illustrate the spatial and temporal dynamics of displacement across the DRC. These examples demonstrate different visualization approaches and provide templates for further analysis and operational use.

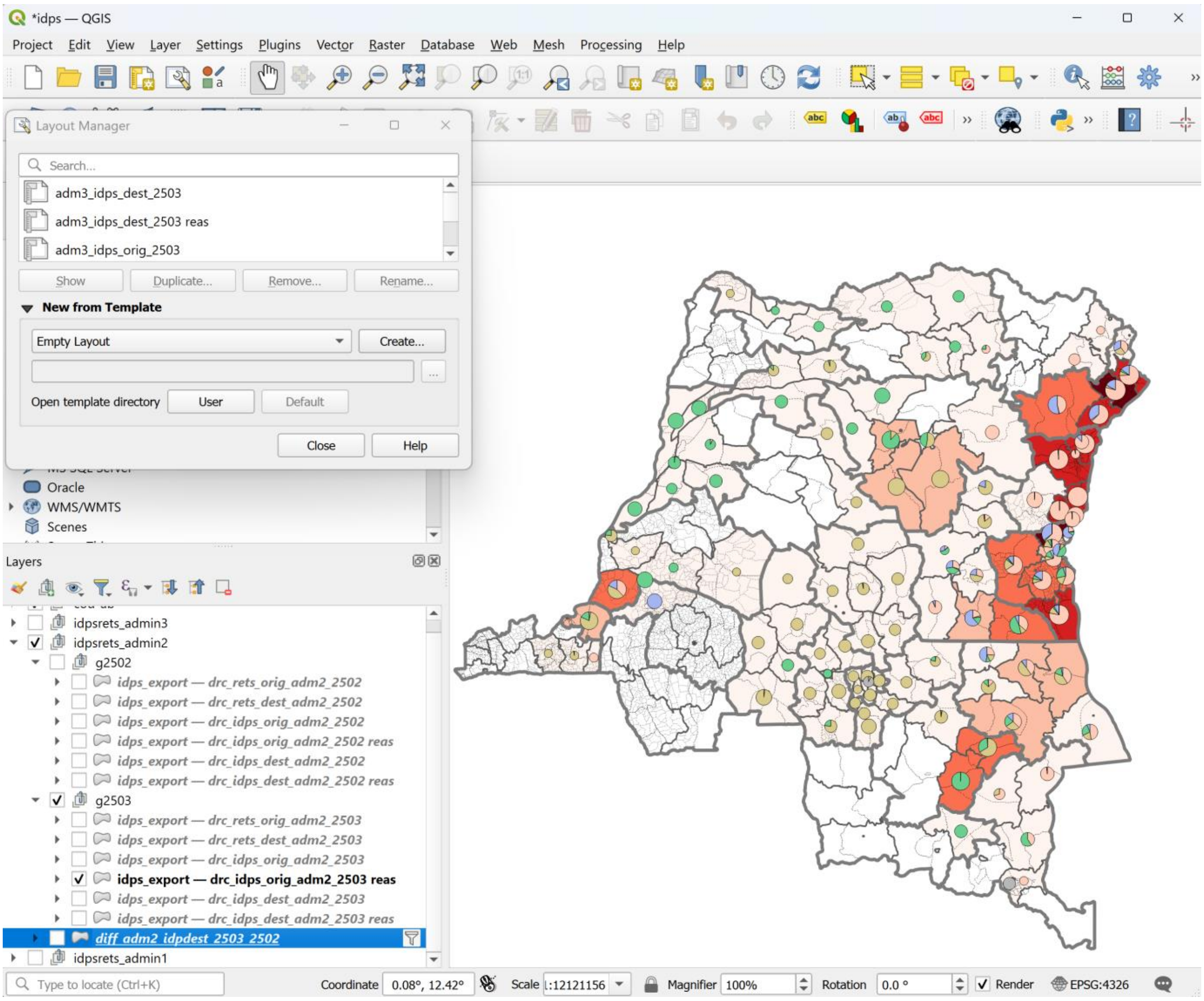

Figure 7 idps.qgz project screenshot comprising multiple layers and layouts. It can be used for IDP/returnee data analysis and as a template for other output layers

The QGIS project contains multiple GIS layers related to IDPs and returnees, together with several prepared print layouts and visualization templates. The project demonstrates different approaches to geovisualizing displacement data using the integrated spatial products generated within this work.

The QGIS project primarily relies on two spatial databases:

- idps.sdb – the main relational SQLite/SpatiaLite database containing all imported administrative boundaries, original displacement tables, linkage tables (lnk_*), metadata, and intermediate processing results. All harmonised geographic features are stored in the el table and include administrative units for both DRC and NGA across multiple levels.

- idps_export.sqlite – a derived SpatiaLite database containing processed GIS-ready displacement layers. These layers were automatically generated through SQL aggregation workflows combining displacement tables (idp_*), linkage tables (lnk_*), and harmonised administrative geometries from idps.sdb.

For desktop GIS analysis, visualization, and cartographic production, we recommend using the layers stored in idps_export.sqlite, as they provide fully prepared and spatially aggregated datasets suitable for direct operational use. For web mapping, lightweight data exchange, and integration with external platforms, the GeoJSON layers stored in the json_export/ directory are generally more suitable.

In addition to per-round GIS layers, the project also includes the Full Wide-Format (FWF) outputs located in the fwf_export/ directory. These datasets provide compact operational layers combining multiple reporting rounds within single GIS files and are particularly suitable for temporal comparison, humanitarian microplanning, and external analytical workflows. Below, we present a selection of representative maps exported from the project.

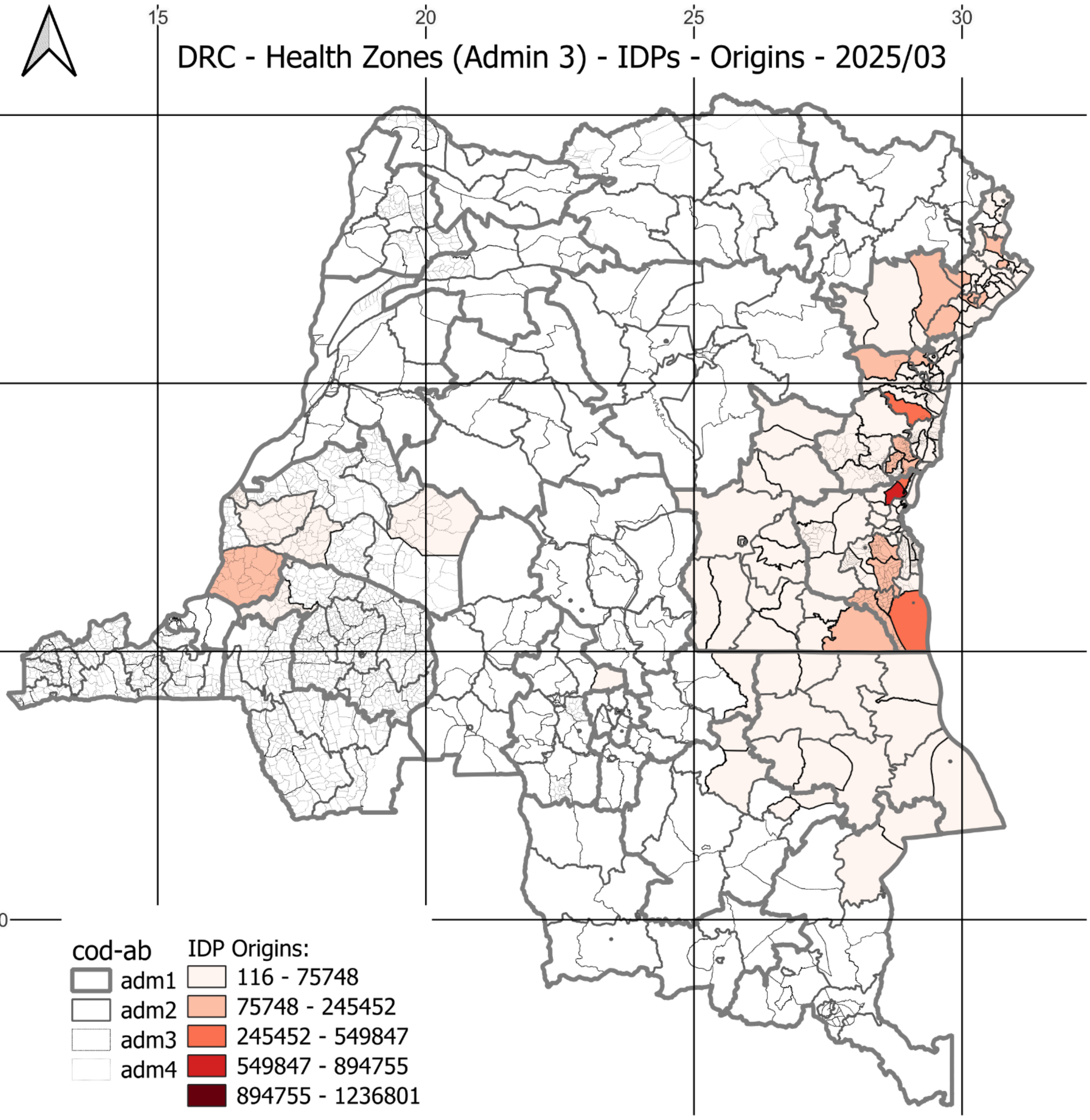

Figure 8 DRC – Health Zones (Admin Level 3) – IDPs' Origins – March 2025. This map shows the health zones in the Democratic Republic of Congo from which internally displaced persons (IDPs) originated in March 2025, based on available displacement records.

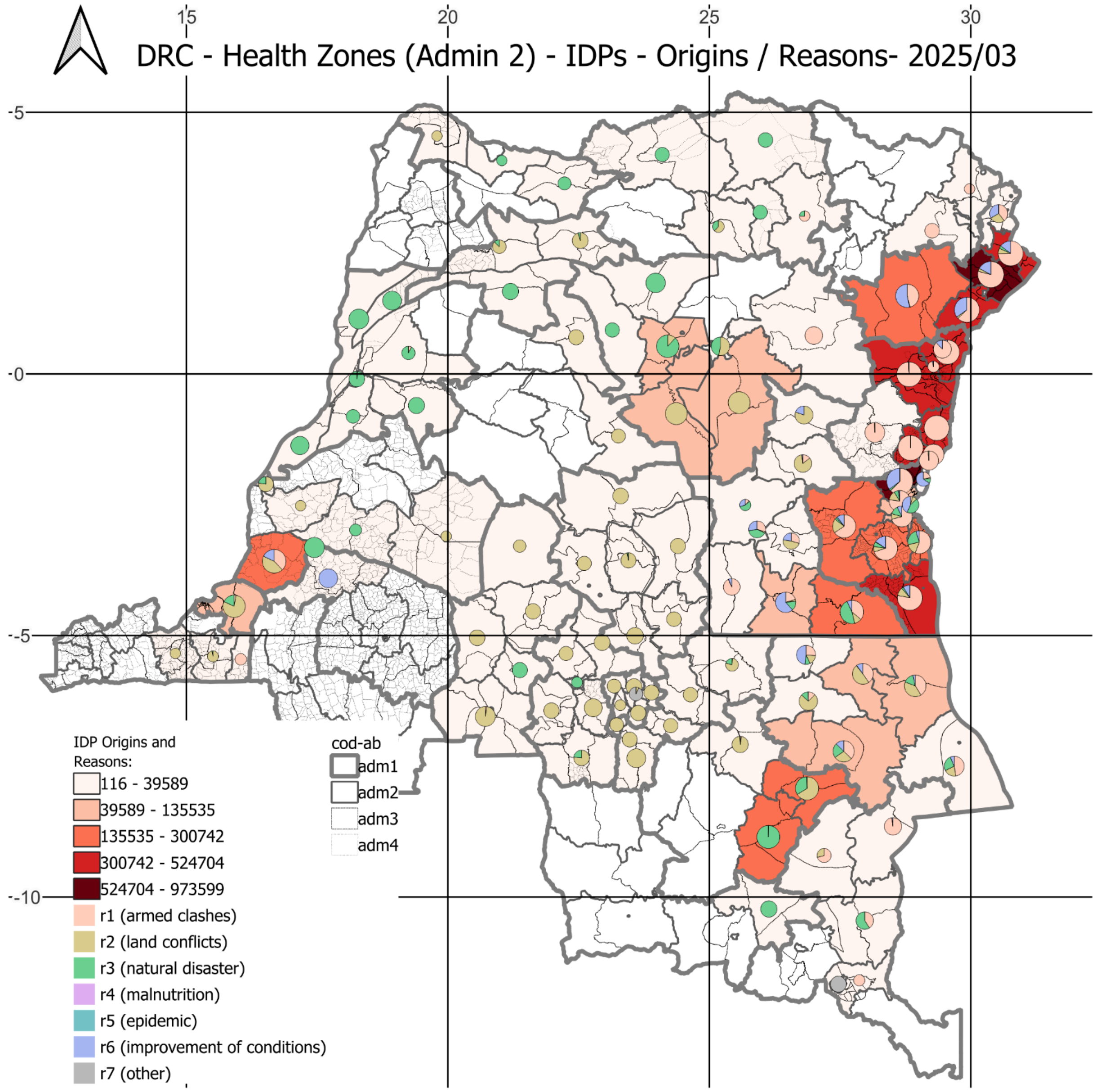

Figure 9 DRC - DRC – Health Zones (Admin Level 3) – IDPs' Origins and Displacement Reasons – March 2025.This map shows the territories (Admin Level 2) in the Democratic Republic of Congo from which IDPs originated in March 2025. It includes information on the reasons for displacement, offering insight into the underlying drivers behind population

movements at the subnational level.

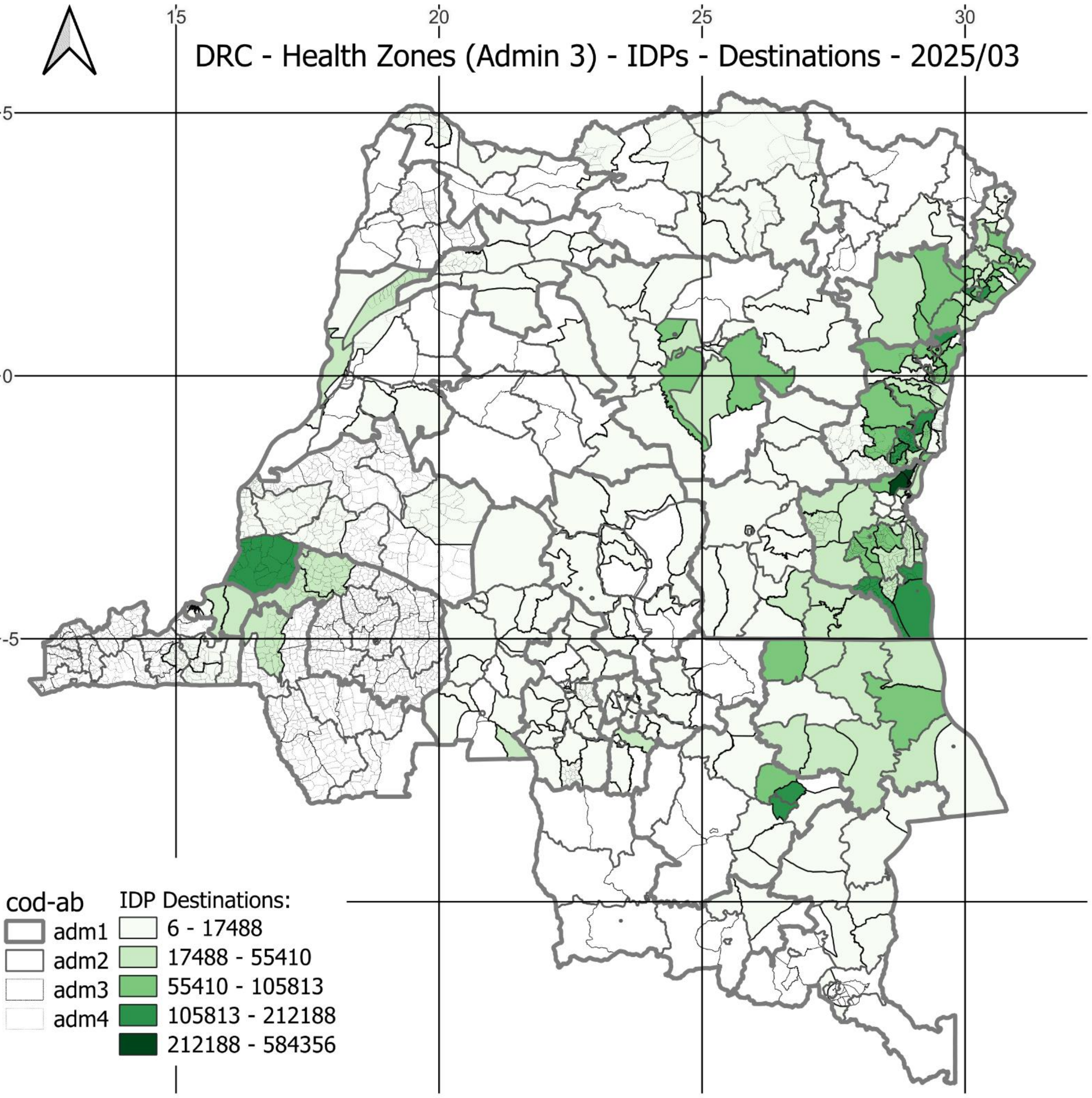

Figure 10 DRC – Health Zones (Admin Level 3) – IDPs' Destinations – March 2025. This map presents the health zones in the Democratic Republic of Congo where internally displaced persons (IDPs) were recorded as arriving in March 2025, illustrating the main destinations of displacement movements during that period.

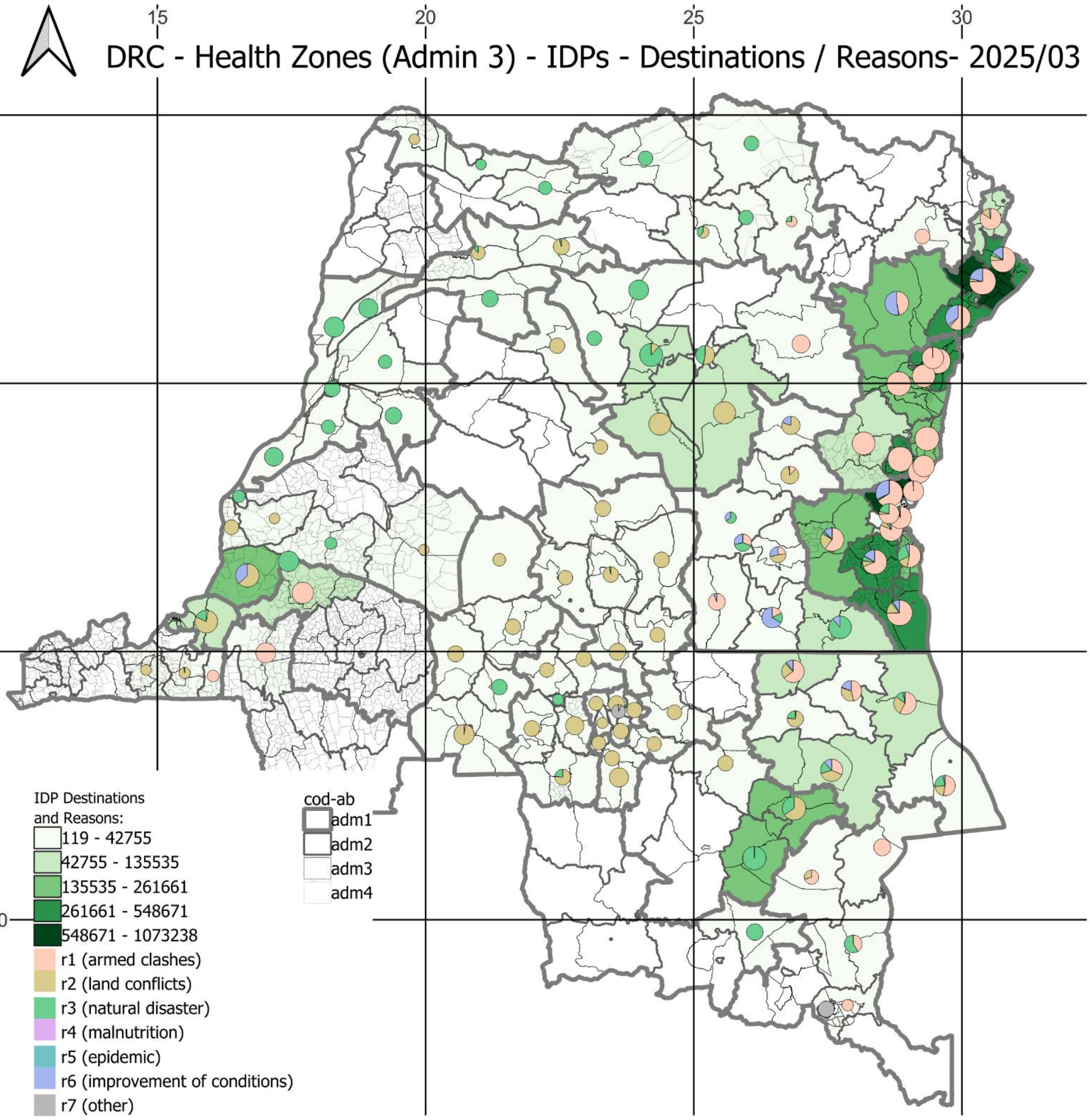

Figure 11 DRC – Health Zones (Admin Level 3) – IDPs' Destinations and Displacement Reasons – March 2025.This map displays the health zones (Admin Level 3) in the Democratic Republic of Congo that received IDPs in March 2025. It also illustrates the reported reasons for displacement, helping to contextualize IDP arrivals and understand the spatial patterns and causes of internal displacement.

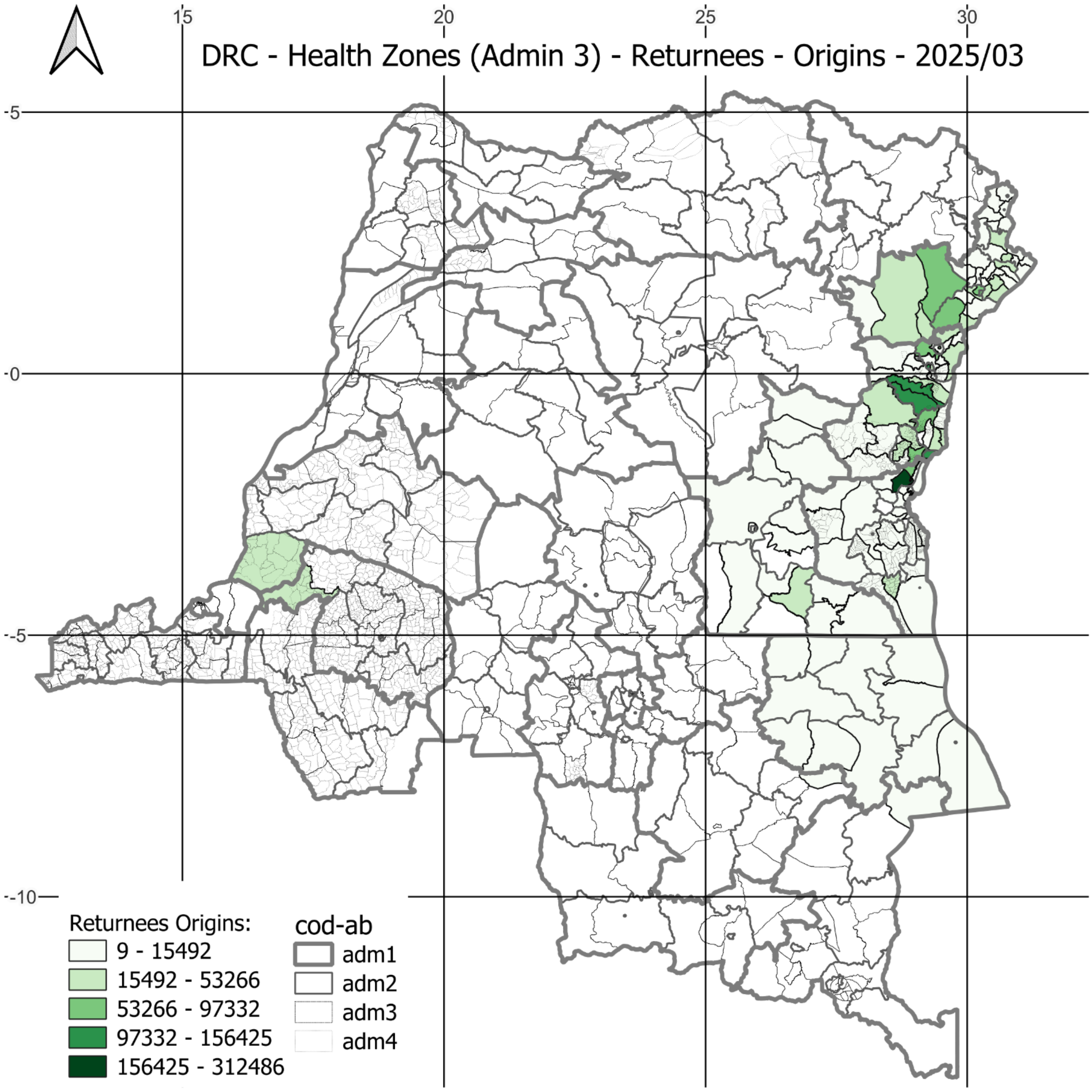

Figure 12DRC – Health Zones (Admin Level 3) – Returnees' Origins – March 2025. This map shows the origin locations of returnees across health zones in the Democratic Republic of Congo for March 2025, based on aggregated displacement data.

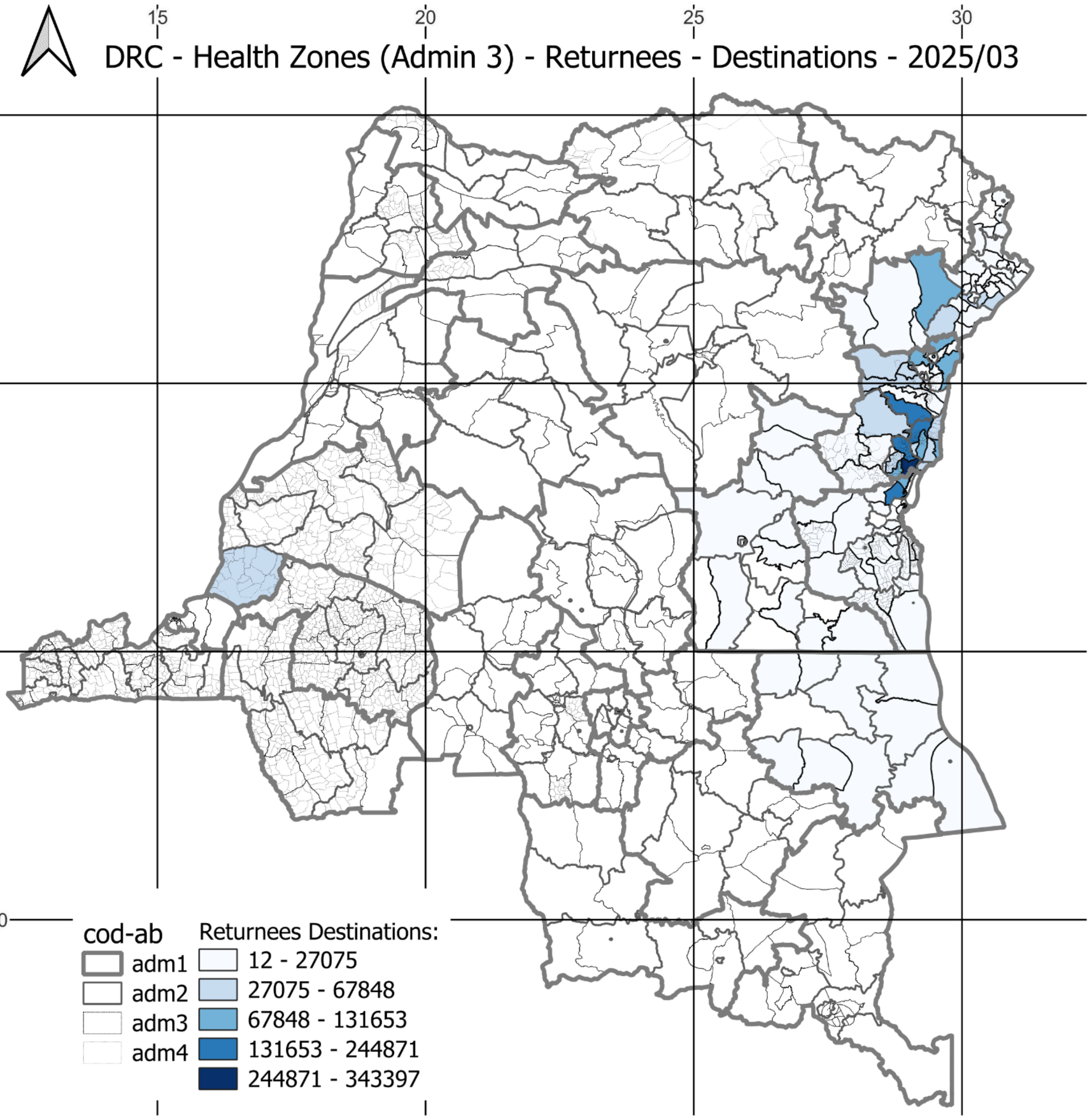

Figure 13 DRC - DRC – Health Zones (Admin Level 3) – Returnees' Destinations – March 2025. This map displays the destination health zones where returnees resettled in the Democratic Republic of Congo during March 2025, as derived from the processed returnee data.

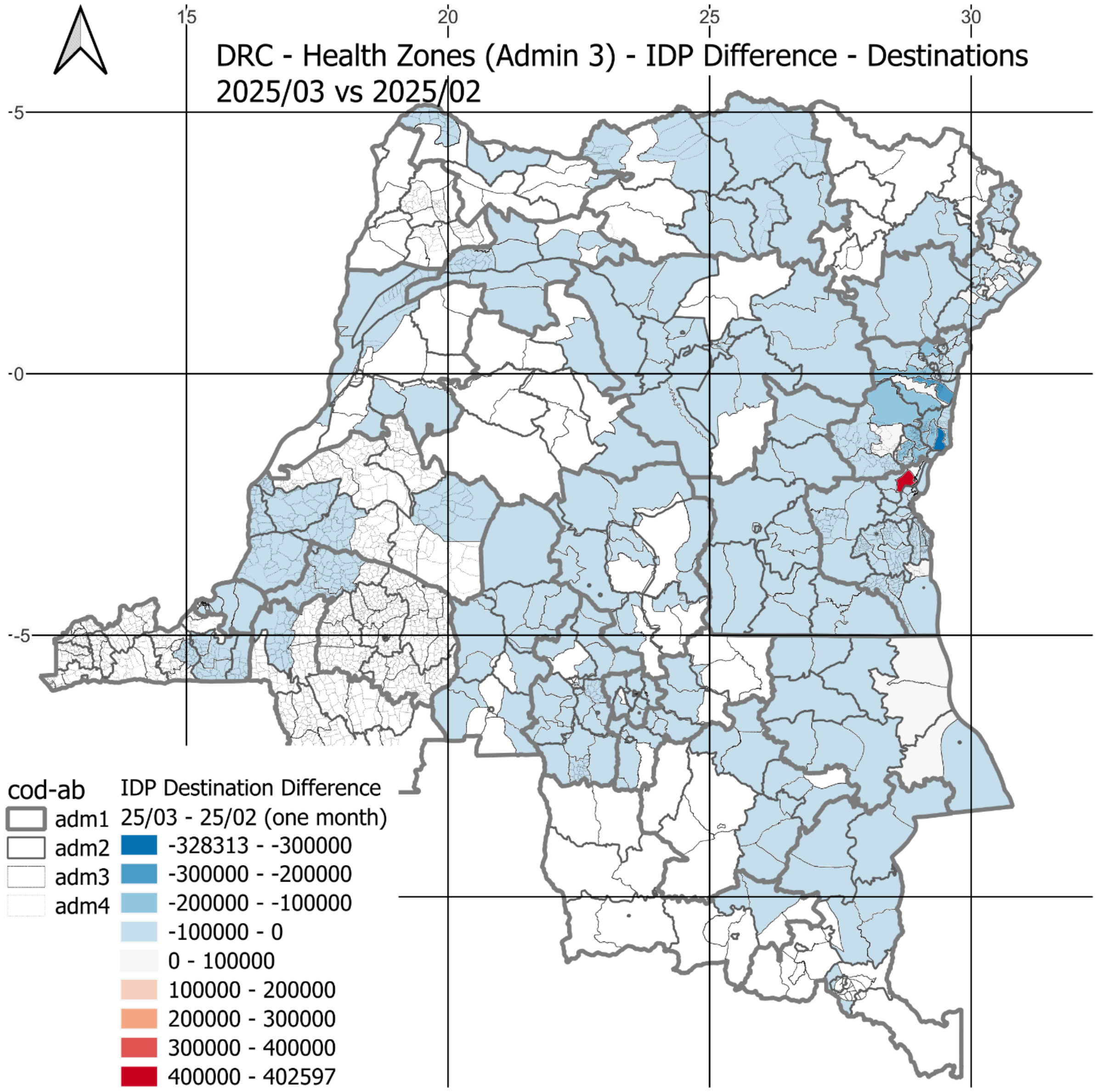

Figure 14 DRC - DRC – Health Zones (Admin Level 3) – IDP Destination Change (March 2025 vs February 2025).This map presents the month-to-month difference in internally displaced persons (IDP) destinations across health zones in the Democratic Republic of Congo, highlighting spatial shifts between February and March 2025.

4. Result Datasets for Nigeria

To generate spatial outputs for Nigeria (NGA), a dedicated Tcl workflow script (https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_nga.tcl&ci=tip) was developed to transform the integrated relational database into GIS-ready operational products. The workflow automatically processes the harmonised SQLite/SpatiaLite database (idps.sdb) and exports spatial outputs in multiple formats, including GeoJSON, GeoPackage, CSV, and SpatiaLite/SQLite.

The workflow initializes a new export database (idps_export.sqlite), enables spatial functionality using the mod_spatialite extension, and attaches the source database containing all integrated displacement and boundary datasets. It then identifies all linkage tables (lnk_*) representing spatial correspondences between displacement records and harmonised administrative boundaries.

The Nigerian workflow additionally performs automated interpretation and normalization of displacement causes across heterogeneous IOM-DTM assessment rounds. Since different assessment rounds use inconsistent textual descriptions for displacement causes, a custom Tcl procedure (reasdict) was implemented to standardize these values into a unified classification system:

1 – Armed banditry / kidnapping

2 – Communal clashes

3 – Farmer–herder conflicts

4 – Insurgency / herdsmen attacks

5 – Natural disasters (e.g., floods)

6 – Other reasons

This normalization enables consistent aggregation and comparison across assessment rounds and regions.

For each reporting round and linkage table, the workflow automatically generates aggregated spatial layers containing:

• Element identifiers (elid, eldid, elname)

• Total numbers of displaced persons or returnees (num)

• Cause-specific displacement counts (c1–c6)

• Polygon or point geometries in EPSG:4326 (WGS 84)

The aggregation is performed using SQL joins between linkage tables (lnk_*), original displacement tables (idp_*), and harmonised administrative boundaries stored in the el table.

The resulting datasets are exported into several output directories:

• json_export/ – GeoJSON files

• gpkg_export/ – GeoPackage files

• csv_export/ – CSV exports

• fwf_export/ – Full Wide-Format (FWF) products

• idps_export.sqlite – combined SpatiaLite database containing all generated GIS layers

In total, 204 GIS layers were generated for Nigeria, representing internally displaced persons (IDPs) and returnee flows across multiple reporting rounds and administrative levels. These outputs include both origins and destinations of displacement events and support analysis at Admin 1–3 levels.

Each exported layer follows a standardized naming convention:

<type>_<direction>_<time>_<admin>_<source>

where:

• type = idp, ret, or idpsrets

• direction = orig or dest

• time = reporting period or assessment round

• admin = administrative level (adm1, adm2, adm3)

• source = spatial linkage approach (hdx, grd3, grd3corr, grd3xy, grd3xynm, or xy)

This convention ensures systematic organization of all displacement layers and compatibility with the integrated GRID3 spatial workflow.

Unlike the DRC workflow, the Nigerian processing pipeline integrates several partially overlapping IOM-DTM assessment systems, including North-East (NE), North-Central/North-West (NCNW), and GTB assessment rounds. As a result, the workflow performs additional harmonisation and on-the-fly merging of multiple datasets originating from different operational regions and reporting structures.

A key operational output of the Nigerian workflow is the generation of Full Wide-Format (FWF) datasets located in the fwf_export/ directory. These represent the primary user-oriented products of this work and provide compact GIS-ready layers combining multiple reporting rounds within single datasets.

For Nigeria, the FWF outputs include:

• adm2* – Local Government Area (LGA) level datasets

• adm3* – Ward-level datasets

The Nigerian FWF products are based on multiple linkage strategies:

• *_grd3xy – coordinate-based aggregation using site GPS coordinates

• *_grd3 – name-matching–based linkage to GRID3 administrative units

• *_grd3xynm – combined coordinate- and name-matching products merging *_grd3xy and *_grd3

The *_grd3xy products provide high spatial accuracy where reliable coordinates are available in the source datasets. However, because coordinates are frequently missing or inconsistent across assessment rounds, the *_grd3xynm products combine coordinate-based estimates with name-matching results to maximize both spatial coverage and geographic precision.

The FWF destination datasets for Nigeria primarily rely on *_grd3xynm products at Admin 3 (ward) level, while origin estimates are primarily derived from *_grd3 linkage products. Admin 2 FWF outputs are subsequently generated through the aggregation of Admin 3 estimates.

The FWF datasets are exported in CSV, GeoJSON, GeoPackage, and SQLite/SpatiaLite formats. Additional *.nogeo.csv files are provided without geometries to facilitate use in spreadsheet software such as Microsoft Excel or LibreOffice Calc.

The generated outputs can be readily visualised and analysed using standard GIS software such as QGIS or ArcGIS. During the development of this project, a dedicated QGIS project (idps.qgz) was prepared internally to support workflow testing, visualization, quality assessment, and production of demonstration maps. However, this QGIS project is not distributed as part of the final public dataset release.

Unlike the DRC case, where most datasets were relatively centralized and structurally consistent, the Nigerian displacement landscape requires integration of several partially overlapping regional IOM-DTM assessment systems. The internal QGIS workflow therefore performed dynamic joins and conditional merging of displacement layers originating from different operational regions and assessment rounds, including North-East and North-Central/North-West datasets.

Although the QGIS project itself is not included in the final release, all required operational outputs are fully provided through the exported GIS layers and databases. For desktop GIS analysis and operational mapping, we recommend using the layers stored in idps_export.sqlite, which contains fully prepared GIS-ready aggregated layers with spatial indexing and metadata. For lightweight web mapping, data exchange, and integration with external systems, the GeoJSON layers located in the json_export/ directory are generally more suitable.

The exported datasets are intended to provide a complete, reproducible, and operationally usable spatial framework for analysing displacement dynamics across Nigeria at multiple administrative levels and across multiple reporting periods.

The following figures illustrate several representative examples of the generated Nigerian displacement products and demonstrate different approaches to visualising the integrated datasets. Together, they highlight the ability of the developed workflow to support both national-scale overview mapping and more detailed subnational analysis of displacement dynamics. The examples include cumulative and round-specific displacement estimates, visualisation of displacement causes, and temporal change analysis between assessment periods. They also demonstrate the practical advantages of the harmonised GRID3-based outputs for operational analysis, humanitarian microplanning, and interpretation of evolving displacement patterns across Nigeria.

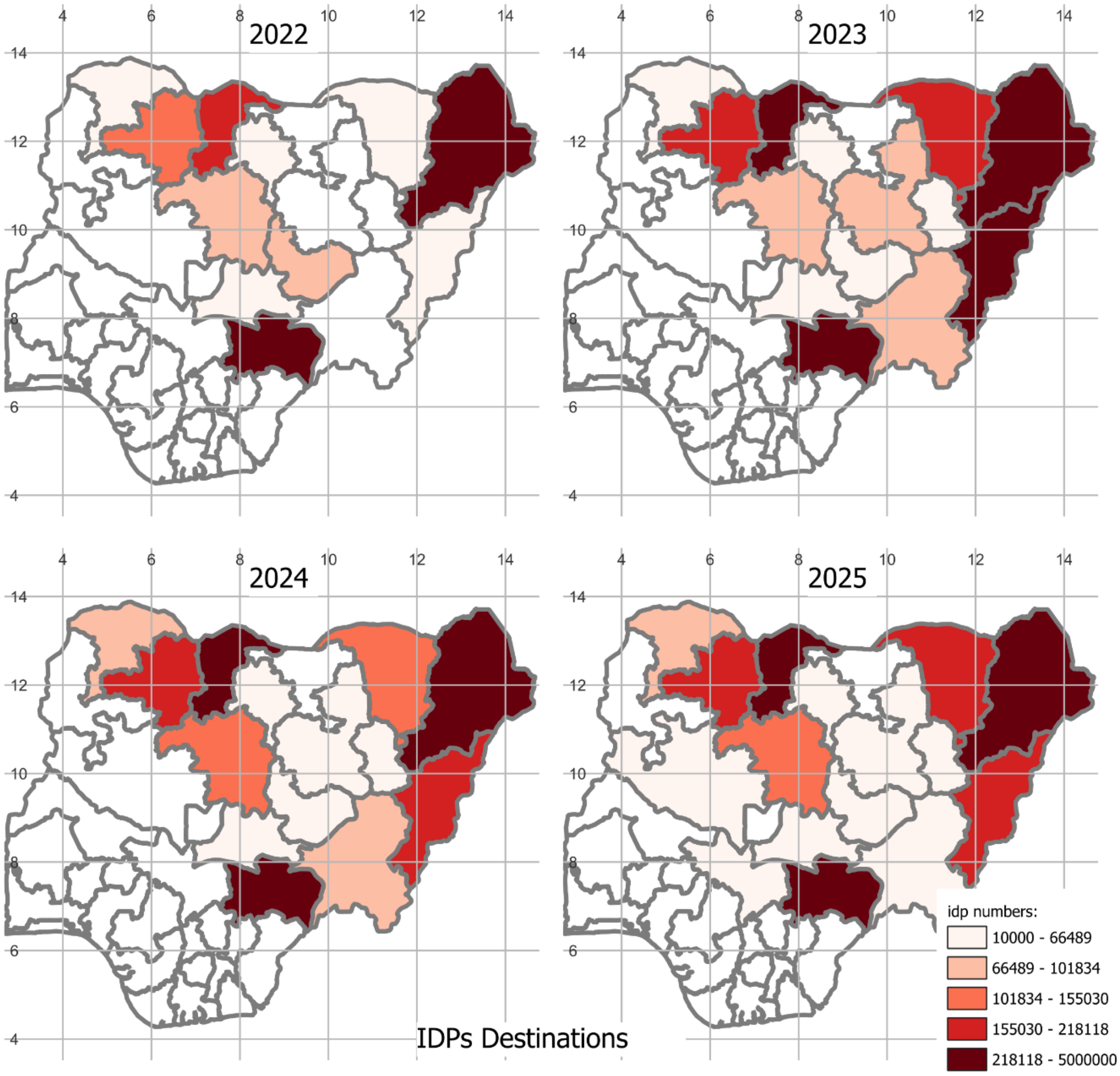

Figure 15 IDP Destinations, 2022-2024, GRID3 Admin 1 Boundaries

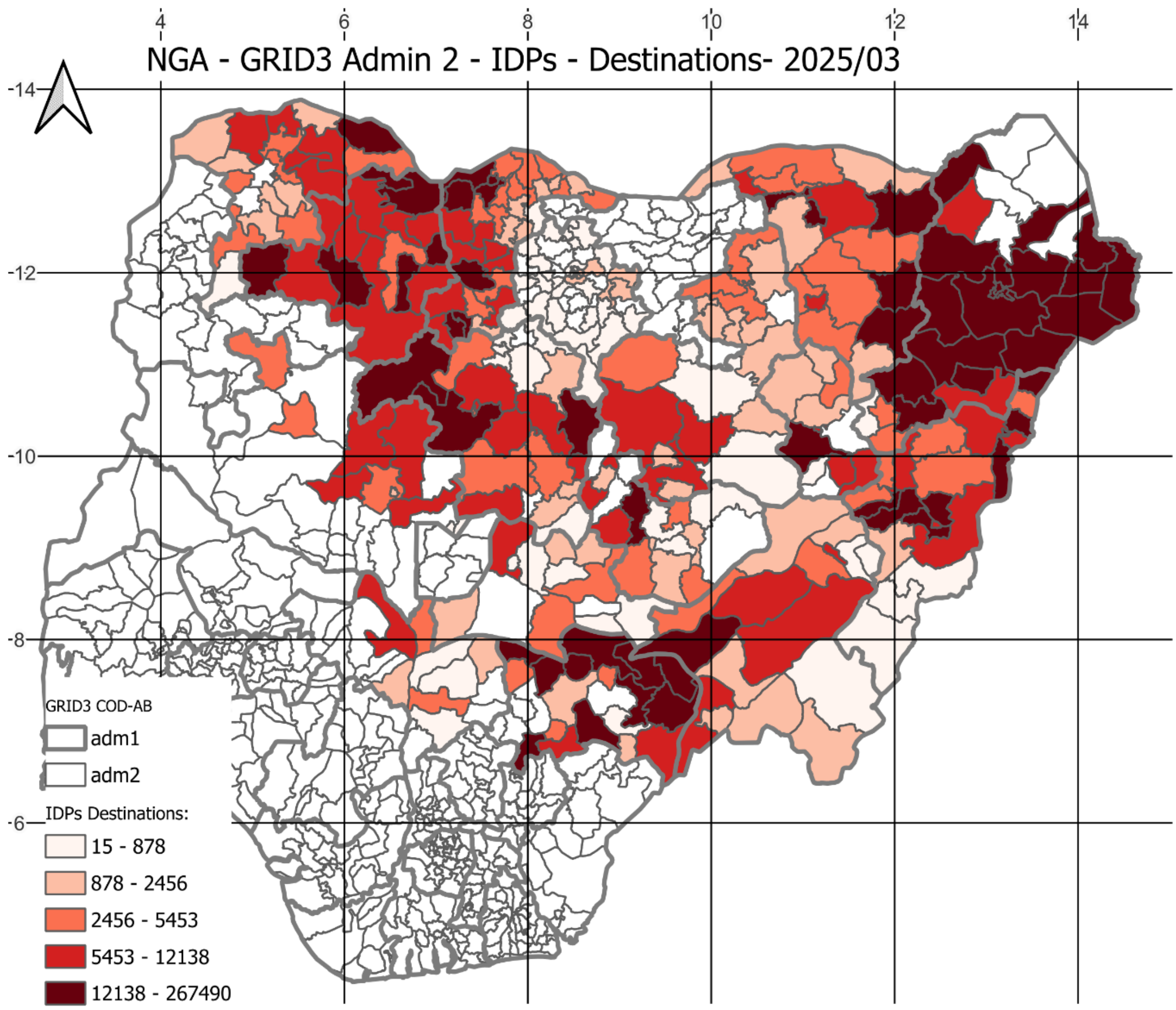

Figure 16 Nigeria – GRID3 Admin Level 2 – IDP Destinations (March 2025). This map shows the spatial distribution of internally displaced persons (IDPs) by destination across Nigeria's Admin 2 units, based on GRID3 boundaries, for March 2025.

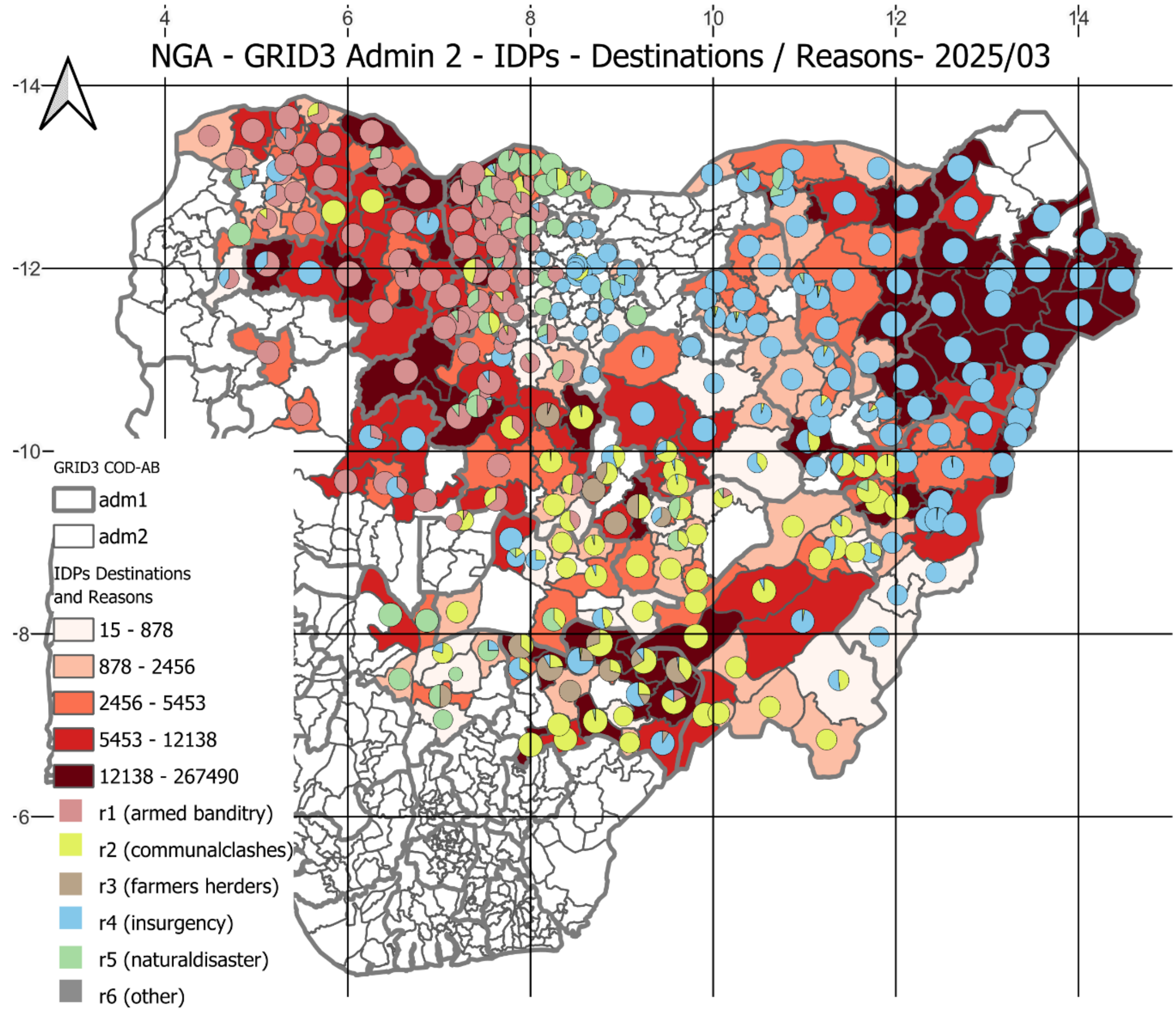

Figure 17 Nigeria – GRID3 Admin Level 2 – IDP Destinations and Reasons (March 2025). This map displays the distribution of internally displaced persons (IDPs) by destination at Admin 2 level in March 2025, with displacement reasons encoded by colours and pie charts.

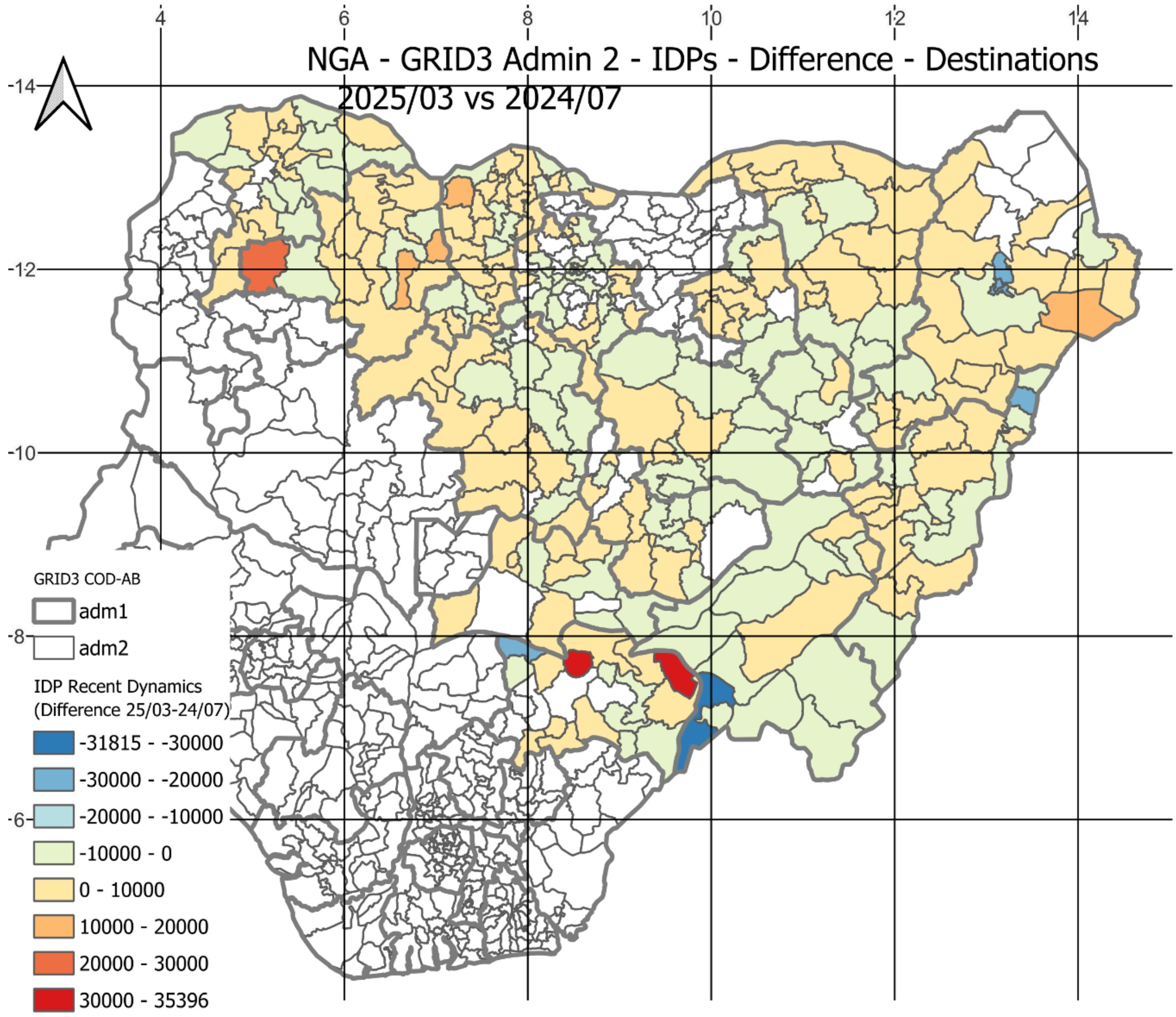

Figure 18 Nigeria – GRID3 Admin Level 2 – IDP Destination Change (March 2025 vs July 2024). This map visualizes the change in the number of internally displaced persons (IDPs) arriving in each Admin 2 unit between March 2025 and July 2024.

5. Quality Assessment

The database assembly, GIS export generation, and quality assessment (QA) workflows were implemented as fully automated procedures within the Tcl-based processing framework. The resulting QA outputs are provided in the nga/QA and cod/QA directories and were generated automatically during execution of the country-specific assembly scripts (sdbasm_nga.tcl and sdbasm_cod.tcl; see PHASE 40 – Clean, Export & QA in https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_cod.tcl&ci=tip and https://igis.n-kov.com/i.htm/file?name=apps/idps/sdbasm_nga.tcl&ci=tip).

The QA framework was designed to evaluate the internal consistency of the generated spatial products and linkage workflows. It performs three complementary categories of checks:

- linkage consistency between displacement tables and spatial boundaries,
- statistical validation of exported GIS layers, and
- nested administrative consistency between different aggregation levels.

In addition to aggregate QA statistics, the workflow also generates detailed positive and negative matching reports (*.pos.psv and *.neg.psv) within the QA/NMS/ directories. These files provide transparent diagnostics of accepted and rejected linkage candidates and support manual inspection of fuzzy name-matching results.

Example entries from the generated QA files are shown below:

```
==> nga/QA/lnk_count.csv <==

tablename,links,rows,equalityindicator,matchingpercentage,nullcellpercentage,consistencyindicator

lnk_idps_dest_2405r12ncnw4sa_adm1_hdx,1637,1637,equal,100,0,consistent

lnk_idps_dest_2206r9ncnw4ba_adm1_hdx,859,859,equal,100,0,consistent

==> nga/QA/nested_num.csv <==

basename,dataB/dataA,idpsOrRets,destOrOrig,admB/admA,min,max,mean,flag

2503r49ne4sa,hdx/hdx,idps,dest,adm1/2,0,133833,36497.0,consistent

2503r49ne4sa,hdx/hdx,idps,orig,adm1/2,0,268771,37056.0,consistent

==> nga/QA/num_gis_stats.csv <==

id,name,count,max,min,sum,csum,geoms,prc,flag

1.i1,idps_dest_2503r49ne4sa_adm1_grd3,6,1704175,52202,2252348,2252348,MULTIPOLYGON-6,100,consistent

1.i2,idps_dest_2503r49ne4sa_adm1_hdx,6,1704175,52202,2252348,2252348,POLYGON-6,100,consistent
```

These example excerpts from the NGA QA outputs demonstrate the structure and purpose of the automated validation framework. The QA procedures evaluate linkage completeness, exported GIS layer consistency, and hierarchical aggregation behaviour across administrative levels. Each QA output file serves a specific validation purpose.

lnk_count.csv

This file evaluates the completeness and consistency of spatial linkages established between displacement tables and administrative boundary layers. For each linkage table (lnk_*), the following indicators are reported:

• the total number of linked records (links) compared with the total number of rows in the original displacement table (rows),

• an equality indicator (equal or nonequal),

• the percentage of successfully matched records (matchingpercentage),

• the percentage of rows containing null or unresolved location values (nullcellpercentage), and

• a consistency flag indicating whether the linkage result requires additional inspection.

Tables are automatically flagged with ATTENTION when a substantial proportion of non-null records cannot be linked to valid spatial units. In the current release, the majority of NGA linkage tables achieved complete or near-complete matching rates with minimal null values, indicating high overall spatial consistency.

num_gis_stats.csv

This file summarizes the principal statistics of all exported GIS layers. For each layer, the following properties are reported:

• the total number of spatial features (count),

• the minimum and maximum IDP/returnee values (min, max),

• the total number of displaced persons (sum),

• the summed values across all displacement-reason columns (csum),

• the distribution of geometry types (geoms), such as POLYGON-6 or MULTIPOLYGON-6, and

• the relative percentage (prc) compared with the largest layer for the same assessment period.

The QA procedure verifies that csum is equal to sum, ensuring consistency between total counts and reason-disaggregated values. Layers are flagged with ATTENTION if substantial reductions in totals occur during spatial aggregation or export procedures. In the final release, all exported GIS layers passed these consistency checks.

nested_num.csv

This file evaluates hierarchical consistency between administrative levels. The procedure compares aggregated IDP totals between nested polygon layers (e.g., Admin 1 vs Admin 2, or Admin 2 vs Admin 3) using spatial containment relationships. For each comparison, the minimum, maximum, and mean differences in aggregated counts are reported.

Cases where lower-level administrative units contain substantially larger totals than their parent polygons are automatically flagged with ATTENTION. This validation step is particularly important for identifying aggregation inconsistencies introduced during boundary harmonisation or correspondence transfers between HDX and GRID3 geometries. In the final release, all nested administrative comparisons were marked as consistent.

All QA metrics were generated automatically during database assembly and export. The procedures systematically evaluate linkage completeness, null-value distributions, aggregation consistency, and nested polygon relationships, and then export the resulting summaries into the country-specific QA/ directories.

All tables and layers flagged with ATTENTION during intermediate processing stages were reviewed manually. In the final release, all identified issues were either corrected or determined to be explainable edge cases related to boundary inconsistencies or unresolved source records.

6. Conclusion

This report presents a comprehensive and reproducible framework for generating spatial estimates of internally displaced persons (IDPs) and returnees in Nigeria (NGA) and the Democratic Republic of the Congo (COD). Using an open-source workflow based on Tcl, SQLite/SpatiaLite, GDAL/OGR, and QGIS, heterogeneous displacement datasets from IOM-DTM, OCHA, HDX, and GRID3 were integrated into a unified spatial database and transformed into harmonised GIS-ready products.

The resulting datasets provide displacement estimates disaggregated by administrative level, displacement direction (origins and destinations), reporting round, and reason for displacement.

For Nigeria, outputs were generated primarily at Admin 2 (LGA) and Admin 3 (ward) levels, while for COD the principal outputs target Health Zones and Health Areas. In total, hundreds of GIS layers were produced and exported in multiple open formats, including GeoJSON, GeoPackage, CSV, and SQLite/SpatiaLite.

A central outcome of this work is the development of harmonized Full Wide-Format (FWF) datasets designed for operational use and rapid analysis. These compact datasets aggregate multiple reporting rounds into unified spatial layers and represent the primary recommended outputs for most users. The FWF products combine multiple geocoding strategies, including p-code matching, constrained fuzzy name matching, and coordinate-based allocation, while preserving compatibility with GRID3 administrative boundaries. For NGA, the highest-quality destination estimates were derived from combined coordinate and name-matching workflows (*_grd3xynm), whereas origins relied primarily on harmonized GRID3 boundary matching (*_grd3). For COD, the principal outputs were generated using harmonized GRID3/HDX correspondence products (*_grd3corr) and direct GRID3 linkage workflows.

The report demonstrates that meaningful spatial analysis of displacement dynamics is possible even under conditions of fragmented humanitarian reporting, inconsistent administrative structures, and incomplete geospatial metadata. Through systematic harmonization, automated QA procedures, and transparent linkage scoring, the workflow enables reproducible spatial analysis across countries, years, and reporting systems.

At the same time, several limitations remain. Displacement datasets continue to vary substantially in structure, terminology, spatial precision, and temporal consistency. Fine-scale administrative matching remains particularly challenging in areas where authoritative boundary datasets are poorly aligned or where reporting units change over time. In COD, for example, correspondence between GRID3 and HDX health-area geometries is often imperfect due to substantial geometric inconsistencies between the underlying datasets. Coordinate information is also unevenly available across assessment rounds and countries.

Nevertheless, this Version 1 release establishes a scalable and extensible foundation for operational displacement mapping and future modelling efforts. The developed tools and workflows are fully reusable and can be adapted to additional countries and displacement contexts. The resulting datasets support humanitarian response, operational microplanning, temporal monitoring, and future high-resolution displacement estimation approaches based on gridded population modelling and spatial covariates.

7. Future Work

The developed framework provides a flexible foundation for future spatial displacement modelling and operational data integration. Several directions for further development have already been identified.

First, future versions of the database will incorporate new displacement assessments and updates for 2026 and beyond. The current workflow was specifically designed to support repeated integration of newly released datasets while preserving compatibility with previous rounds and maintaining harmonised outputs across administrative levels.

Second, additional work is planned to improve the matching percentage at fine administrative levels, particularly for Nigeria wards and DRC health zones and health areas. The current implementation intentionally uses conservative score thresholds (score 2 for Nigeria and score 1 for DRC) to minimise false-positive matches. Future versions will additionally produce outputs generated using more permissive matching thresholds (e.g., scores 3 and 4 instead of the current scores 1 and 2), followed by AI-assisted filtering and validation of potentially incorrect matches. This approach may significantly increase coverage while preserving acceptable matching quality.

Third, future work will focus on improving automated interpretation of complex IOM-DTM spreadsheets. Source displacement datasets frequently contain inconsistent column names, changing table structures, merged cells, and varying reporting conventions between rounds and countries. AI-assisted approaches may substantially improve automatic recognition of displacement indicators, demographic variables, origin–destination relationships, and geographic references from heterogeneous spreadsheets.

For the DRC, future releases will increasingly rely directly on IOM-DTM source data instead of OCHA-derived intermediate products. This transition is expected to improve consistency of origin–destination relationships, preserve more detailed metadata, and allow extraction of additional demographic information where available.

Additional future directions include improving fuzzy name matching using expanded locality dictionaries, OpenStreetMap and other external gazetteers, and AI-assisted verification of ambiguous geographic correspondences. The developed framework may also support future child-focused displacement analysis, as age-disaggregated information available in original source datasets is preserved within the integrated database and can be extracted for subsequent demographic modelling.

Overall, the project demonstrates the need for continued development of reproducible and transparent spatial harmonisation workflows capable of transforming fragmented displacement data into operational geospatial evidence for humanitarian analysis and microplanning

8. Data and Software Availability

Software: https://igis.n-kov.com/i.htm/dir?ci=tip&name=c and https://igis.n-kov.com/i.htm/dir?ci=tip&name=apps/idps (MIT License: https://opensource.org/licenses/MIT) DOI: 10.5281/zenodo.15794295

Entire Data Archive: https://ar.n-kov.com/p/idps.zip (Creative Commons Attribution 4.0 International Creative Commons Attribution 4.0 International) DOI: 10.5258/SOTON/WP00826

User-friendly FWF extraction: https://ar.n-kov.com/p/idps_cod_nga.fwf.zip See README.txt files in the nga and cod subdirectories (Creative Commons Attribution 4.0 International Creative Commons Attribution 4.0 International)

Entire Data Archive description (also included in idps.zip/idps/README.txt):

========== README.txt ==========

Spatial estimates of displaced persons,

Nigeria (NGA) and Dem. Rep. Congo (COD),

version 1

-------- Overview --------

This directory contains the integrated spatial

displacement dataset (version 1) for the Democratic Republic of the

Congo (COD) and Nigeria (NGA), prepared as part of the GRID3 – Phase 2

Scaling initiative. The data include harmonised IDP and returnee

datasets, relevant administrative boundaries, geocoded displacement

summaries, and a set of derived spatial layers in CSV, SQLite,

GeoPackage, and GeoJSON formats.

For NGA, '2' is the maximal accepted score for name matching; for COD,

'1' is the maximal accepted score for name matching.

---------- Directory Structure ---

2 directories for countries:

cod/* - All files related to the Democratic Republic of the

Congo (COD).

nga/* - All files related to Nigeria (NGA).

(main)

a. `idps.sdb` - the main relational database comprising all

original data, results, and metadata (recommended for advanced

users).

(exported GIS IDP layers - recommended for regular users)

b. `idps_export.sqlite`

– SpatiaLite exports of integrated datasets.

c. `csv_export/` – Exported CSV files with

destination and origin estimates by administrative level,

period, and displacement type.

d. `gpkg_export/` – Exported GeoPackage files with

destination and origin estimates by administrative level, period,

and displacement type.

e. `json_export/` – Exported GeoJSON files with

destination and origin estimates by administrative level, period,

and displacement type.

f. 'fwf_export' - Exported Full Wide-Format (FWF) GeoPackage, GeoJSON,

CSV, and Spatialite/SQLite data. It is the most user-friendly way

to use this dataset, but it provides simplified information. This a a compact data format that stores all rounds of assessment in columns and does not include reasons of IDPs. Notice, *.nogeo.csv do not include geometries; that make it convenient to use in Excel/LibreofficeCalc. To generate this data the following sources were used (they represent data with the highest percentage and accuracy of allocated IDPs):

- COD Admin 4 (Health Areas) - *_grd3,
- COD Admin 3 (Health Zones) - aggregated Admin 4,
- NGA Admin 3 (Wards) - *_grd3xynm/*_grd3 for destinations (_grd3xynm where available), *_grd3 for origins
- NGA Admin 2 (LGAs) - aggregated *_grd3xynm/*_grd3 Admin 3 (wards) layers for destinations and *_grd3 for origins.

Notice, *.sqlite data comprise all GIS layers in a single file.

(original data from external sources)

f. `cod/hdx/` – HDX boundary shapefiles at admin levels 1–3, plus health zones/areas.

g. `cod/grd3/` – GRID3 administrative boundaries (mainly in GeoPackage format).

h. `idp/` – Cleaned and aggregated IDP and returnee data from IOM-DTM and HDX (CSV + Excel).

(quality assessment and logs)

i. `log/*` - logs: main.txt - main log, ph*.txt - phase logs, found_conns_adm/idps*.tmp.psv - found matching logs (pipe-separated values);

j. `QA/*` - quality assessment: lnk_count.csv - link count statistics, nested_num.csv - nested IDP number statistics, num_gis_stats.csv - statistics derived from exported GIS layers;

k. `QA/corrs_*.csv` - identified HDX/GRID3 boundary correspondences derived using a geometric approach;

l. `QA/NMS/lnk*.psv` - reports of best IDP row–geometry correspondences with scores:

lnk*.pos.psv - positive (accepted) matches, lnk*.neg.psv - negative (rejected) best-matching pairs.

m. `QA/fwf_stat*.csv` - reports the difference in IDP numbers in the generated fwf-layers at highest admin level with the following corresponding *_grd3 GIS layers.

-------- File Naming Conventions -----

All IDP/Returnee (export) files (tables) follow the convention:

TYPE_DIRECTION_MRD_ADM_GEO

TYPE - data type: either 'idp' (IDPs) or 'ret' (returnees).

DIRECTION - either 'orig' (origin) or 'dest' (destination).

MRD - month/round/description:

months in YYMM format, round of assessment

(rN, where N is the round number), and description

(either 'ba' - baseline assessment or 'sa' - site assessment).

BND - source of geographic information:

- 'grd3' - GRID3,
- 'hdx' - HDX (OSGOF for NGA, OCHA for COD),
- 'grd3corr' - GRID3 boundaries where IDP numbers were transferred from an '_hdx' dataset using geometric matching,
- 'xy' - GPS coordinates from IDP spreadsheets,
- 'grd3xy' - GRID3 boundaries aggregating 'xy' data within their administrative units,
- 'grd3xynm' - GRID3 boundaries produced by merging '_grd3' and '_grd3xy'.

All exported GIS layers (CSV, GeoPackage, GeoJSON, SQLite) and tables in idps.sdb follow this convention (linking tables have the prefix 'lnk').

-------- File Formats --------

- `.csv` – Comma-separated value tables of displacement figures by location and reason.
- `.xlsx` – Excel versions of data inputs.
- `.gpkg` / `.geojson` – Geospatial layers in open vector formats.

- `.shp` – ESRI Shapefiles for administrative

boundaries.

- `.sqlite` - SQLite/SpatiaLite database comprising all exported GIS

  layers in a single file.

- `.sdb` - SQLite/SpatiaLite relational database containing all

  original data, results, and extended metadata.

--------- Data model of the exported data ------

Each exported GIS file (layer) contains the following columns:

- 'elid' - (INTEGER) unique ID of geometric features (copied from

  the idps.sdb 'el' table);

- 'eldid' - (TEXT) unique ID copied from the boundary source data

  (GRID3 or HDX);

- 'elname' - (TEXT) name of the administrative unit;

- 'eone' - (TEXT) first extra column copied from the boundary source

  data (see the following section);

- 'etwo' - (TEXT) second extra column copied from the boundary source

  data (see the following section);

- 'num' - (INTEGER) number of IDPs/Returnees;

- 'cN' - (INTEGER) number of IDPs/Returnees caused by reason N:

NGA reasons - 1 (armed banditry), 2 (communal clashes), 3 (farmer–

herder conflict), 4 (insurgency), 5 (natural disasters), and

6 (others),

COD reasons - 1 (armed clashes), 2 (land conflicts),

3 (natural disasters), 4 (malnutrition), 5 (epidemics),

6 (improvement of conditions), and 7 (others).

--------- EONE and ETWO content description ----

The following table shows the content of the 'eone' and 'etwo'

columns (copied from original boundary files).

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%ADMIN_LEVEL,    EONE,       ETWO       %
%%%%%%%%%%%% Nigeria GRID3 %%%%%%%%%%%%%%%%%%%%%%%%%%%
%ADMIN 1,      statecode,    NULL       %
```

%ADMIN2, statecode, statename %

%ADMIN3, lganame, statename %

%%%%%%%%%%%% Nigeria HDX %%%%%%%%%%%%%%%%%%%%%%%%%%

%ADMIN 1, NULL, NULL %

%ADMIN2, ADM1_PCODE, ADM1_EN (name)%

%ADMIN3, ADM2_EN (name), ADM1_EN (name)%

%%%%%%%%%%%%% DRC GRID3 %%%%%%%%%%%%%%%%%%%%%%%%%%

%ADMIN3 (HZ), antenne, province %

%ADMIN4 (HA), zonesante, antenne %

%%%%%%%%%%%%% DRC HDX %%%%%%%%%%%%%%%%%%%%%%%%%

%ADMIN 1, NULL, NULL %

%ADMIN2, ADM1_PCODE, ADM1_FR (name)%

%ADMIN3 (HZ) PROVINCE, NULL %

%ADMIN4 (HA) ZS, Province %

%%%
%

--------- Data model of `idps.sdb` ------------

`idps.sdb` is a relational database intended for advanced users. It is the main holistic database of the project. It comprises all original data, calculations, results, and extensive metadata.

The following tables are provided:

- 'el' - all imported administrative boundaries (imported with geo2splite.tcl), consisting of the following columns: 'id', 'did' (ID copied from the original boundary layer), 'type' (geometry type), 'lod' (level of detail / admin level), 'name', 'eone' (extra column one), 'etwo' (extra column two), 'dsrc' (data source), 'geom' (geometry), and 'garea' (bounding-box area of the geometry);
- 'elview' - a view based on the 'el' table implemented for technical purposes (to enable data export to GRASS GIS);
- 'idp_*' - IDP/Returnee data tables;

- 'lnk_*' - tables linking 'idp_*' tables and 'el', consisting of the following columns: 'lid' ('el' ID), 'rid' (IDP rowid), 'ctype'

(connection type), and 'score' (matching score);

- 'meta_cols' - metadata describing columns of IDP tables:

tname (table name), cname (column name), type (column type),

nullprc (percentage of NULL values), length, min (minimum),

and max (maximum);

- 'meta_tabs' - metadata describing IDP tables:

tname (table name), fname (file name), fsize (file size in bytes),

ncols (number of columns), and nrows (number of rows).

For more details see the following report:

Spatial estimates of displaced persons,

Nigeria (NGA) and Dem. Rep. Congo (COD),

version 1

Acknowledgement

This work was part of the GRID3 – Phase 2 Scaling project, with funding from the Bill & Melinda Gates Foundation (INV-044979). Project partners included GRID3, the Center for International Earth Science Information Network (CIESIN) at the Columbia Climate School at Columbia University and WorldPop at the University of Southampton.

ANNEX A. Detailed Installation Instructions and Data Processing Guide

(command instructions start with '$', tcl command shell starts with '%')

Dependencies and Installation

The software was designed for a Unix environment and can be run seamlessly on GNU/Linux, *BSD (including macOS), and Windows (via WSL). It was developed and intensively tested on Windows WSL (with a standard Debian image), Ubuntu, and Debian GNU/Linux. Please install

Windows Subsystem for Linux (WSL)

If you use Windows, we recommend installing Debian GNU/Linux under WSL. Use the following instructions:

https://wiki.debian.org/InstallingDebianOn/Microsoft/Windows/SubsystemForLinux

https://learn.microsoft.com/en-us/windows/wsl/install

Open PowerShell application

PowerShell instructions - Debian example :

$ wsl --install -d Debian

PowerShell instructions - Ubuntu example :

$ wsl --install -d Ubuntu

After setting up the login details, we are already in Debian. If you want to exit Debian, type 'exit'. If you need to start Debian, simply type 'wsl'.

Installing Global Dependencies (Debian-based distros)

For Ubuntu run this (not for Debian):

$ sudo add-apt-repository universe
$ sudo add-apt-repository multiverse
$ sudo apt update

$ sudo apt upgrade

Log in to your system and run:

$ sudo apt update ; sudo apt upgrade

$ sudo apt install tcl tcl8.6 tcllib build-essential sqlite3 libsqlite3-dev libsqlite3-tcl libsqlite3-mod-spatialite spatialite-gui spatialite-bin gdal-bin grass unzip tcl-thread

Type password

Install libsqlitefunctions:

$ sudo apt install wget

$ cd

$ wget -O extension-functions.c https://sqlite.org/contrib/download/extension-functions.c?get=25

$ gcc -fPIC -shared extension-functions.c -o libsqlitefunctions.so -lm

$ sudo cp libsqlitefunctions.so /usr/local/lib/

$ sudo ldconfig

Installing the Developed Software

The following software should be installed: tcl, tcllib, build-essential, sqlite3, tcl-sqlite3, unzip, tcl-thread, libspatialite (mod-spatialite), gdal, and grass.

N-kov's Libs and Tools:

$ wget https://www.n-kov.com/LibsAndTools/lat.tar.gz

$ tar xvzf lat.tar.gz

$ cd LibsAndTools/

$ sudo make

Test:

$ tclsh

% package require n-kov

If it returns the version number, it works fine!

% exit

Installing Tiles.CF:

$ cd GNUMakeTcl

$ wget https://tiles.n-kov.com/i.htm/tarball/4f7f48324d/Tiles.CF-4f7f48324d.tar.gz

$ tar xvzf Tiles.CF-4f7f48324d.tar.gz

$ mv Tiles.CF-4f7f48324d/ tilescf

$ chmod +x tclshc

$ make NAME=tilescf

$ cd /tmp/tilescf

$ sudo make PREFIX=/usr/lib

Test:

$ tclsh

% package require grs

If it returns 0.4, it works fine!

% exit

Installing IGIS.TK:

Download the latest check-in tar.gz or zip archive and extract from https://igis.n-kov.com/i.htm/timeline

Currently, the latest version: 2026-05-19 16:04 RC4c NGA and DRC Leaf check-in: 4def73a29b user: stn tags: trunk. Click n this. It will open up a new window.

Right click on 'Tarball' and 'Copy link'. Update the 'wget' command link with this link.

The IGIS commands will also need an update. Just enter tar xvzf IGIS in the command prpmt and hit TAB. It ill fill in the needed info.

E.g.:

$ sudo mkdir /prs

$ sudo chown mydeb:mydeb /prs

$ cd /prs

$ wget https://igis.n-kov.com/i.htm/tarball/4def73a29b/IGIS.TK-4def73a29b.tar.gz

$ tar xvzf IGIS.TK-4def73a29b.tar.gz

$ mv IGIS.TK-4def73a29b igistk

Running Scripts (first time)

$ cd

$ export W=$(pwd)

$ echo 'export W=$HOME' >> ~/.bashrc

$ cd $W

$ wget https://ar.n-kov.com/p/idps.ph00init.zip

(Note: The zip file contains all IDP information, boundaries, structure, etc, thus all inputs before doing the actual calculations)

$ unzip idps.ph00init.zip

The following command will print the instructions for how to run the code (read them carefully):

$ tclsh /prs/igistk/apps/idps/sdbasm_nga.tcl

Example for Nigeria

After everything is installed:

Open PowerShell

Type ‘wsl’

Then run the country command:

$ rm -rf ~/idps ; cp -r $W/idps.ph00init/idps ~/ ; cd ~/idps ; tclsh /prs/igistk/apps/idps/sdbasm_nga.tcl BATCH > main.nga.txt 2>&1 ; mv main.nga.txt nga/logs/main.txt

To monitor the execution of the commands, you need to open another NEW session with wsl and run this code: tail -f ~/idps/main.nga.txt

Example for DRC

The following command will print the instructions for how to run the code (read them carefully):

$ tclsh /prs/igistk/apps/idps/sdbasm_cod.tcl

for Dem. Rep. of the Congo (COD) - to be run after NGA country driver:

(rm -rf ~/idps/cod; mkdir ~/idps; cp -r $W/idps.ph00init/idps/cod ~/idps/ ; cd ~/idps ; tclsh /prs/igistk/apps/idps/sdbasm_cod.tcl BATCH > main.cod.txt 2>&1 ; mv main.cod.txt cod/logs/main.txt)

to monitor the execution progress use 'tail -f ~/idps/main.cod.txt' in another terminal session

The results will be saved in the (Linux) Debian Home directory:

\\wsl.localhost\Debian\home\mydeb\idps

GRID3 Package Preparation – FWF Extraction – Final Step

Run:

$ tclsh /prs/igistk/apps/idps/fwfpack.tcl

It creates 'idps_cod_nga' in the home directory, comprising FWF extractions and relevant README.txt files in the nga and cod subdirectories. Also, it creates 'idps_cod_nga.fwf.zip' in the home directory final archive.

Updating the database

To make the data actual, download the recent XMLS from IOM-DTM https://dtm.iom.int/datasets for Nigeria and HDX https://data.humdata.org/ for DRC. Rename it according to the name conventions described in README.txt (e.g., idps_2201r8ncnw4ba.xlsx), save it to ~/idps/nga/idp/xls or ~/idps/cod/idp/xls, save as CSV to ~/idps/nga/idp/, and re-run tclsh igistk/apps/idps/sdbasm_nga.tcl and tclsh igistk/apps/idps/sdbasm_nga.tcl.