

SUBNATIONAL POPULATION ESTIMATES COMPONENTS FOR THE HIV SPATIAL DATA REPOSITORY

Robert M. Leddy, Jr.
Geographic Studies Branch
Population Division
U.S. Census Bureau

January 29, 2008
Revised August 30, 2011

Contents in This Document

1. Subnational Boundary Files with Population Data for the HIV Spatial Data Repository
 2. Data in a Country Shapefile and Spreadsheet
 3. Population Estimates/Projections 2000-2015
 4. Shapefiles
 5. Spreadsheets
 6. Metadata
 7. Standards
 8. Updates, Upgrades, Additions
- Data Dictionary

1. Subnational Boundary Files with Population Data for the HIV Spatial Data Repository

The U.S. Census Bureau has uploaded subnational 5-year age/sex group population estimates/projections for 2000-2015 for designated countries in the President's Emergency Plan for AIDS Relief (PEPFAR) program on the HIV Spatial Data Repository. There are **3 components**:

- The **shapefile** (ArcGIS) of the country's subnational areas with the age/sex group population data in the attribute table
- A stand-alone **spreadsheet** (Excel) with the subnational population data
- A **metadata** file in a U.S. Federal Geographic Data Committee-compliant "htm" format exported from the shapefile

Files of the individual countries comprise data at the lowest administrative level (first, second, or third) for which both census data and digital boundary shapefiles are available.

2. Data in a Country Shapefile and Spreadsheet

- **Subnational Division Names.** These consist of the names of the subnational divisions of the country down to the lowest available level (first, second, or third).

Names and spellings comply with the U.S. Board on Geographic Names (BGN) standards. When BGN spellings are not available, spellings from official census data are used.

- **Subnational Population Estimates/Projections.** These are in 5-year age/sex groups up through 80 years and older for the years 2000-2015.
- **Subnational Area Codes.** Based on availability, these consist of the U.S. Government Federal Information Processing System (FIPS) codes (at the country and first-administrative division levels only), official statistical area codes, the World Health Organization *Second Administrative Level Boundary (SALB)* codes, and any other commonly used codes.
- **Comment.** Usually of alternate names or spellings or relevant historical information.
- **Linkage: Spreadsheet to Shapefile.** The data from the spreadsheet are joined to the shapefile attribute table through the common field "GEO_MATCH," consisting of a unique concatenation of administrative names for each administrative area.

A **Data Dictionary** describing the shapefile/spreadsheet fields in more detail is appended at the end of this document. It also is provided in a separate worksheet in each country spreadsheet.

3. Population Estimates/Projections 2000-2015

- Population estimates/projections are provided by **5-year age/sex groups** up through 80 years and older for each subnational area for 2000-2015.
- The estimates/projections are based on population data from the country's most recent available **census**.
- With a series of **raking models**, the age/sex data derived from the census are made consistent with the country's **national midyear 2000-2015 5-year age/sex group projections** in the U.S. Census Bureau's ***International Data Base*** (<http://www.census.gov/ipc/www/idb/>).
- A description of the **raking procedures** is provided in the separate document "Methods for Estimating and Projecting Population by Age and Sex for Subnational Areas of Countries in the HIV Spatial Data Repository."

4. Shapefiles

- **Attribute Data.** The attribute data of a country's shapefile comprise the country's subnational 5-year age/sex group population estimates/projections for the years 2000-2015 along with subnational area names, codes, and comments described in Item 2. The data are linked from the spreadsheet using the "GEO_MATCH" field.
- **U.S. Government Standards.** The subnational boundaries within a shapefile conform to areas as recognized by the U.S. Government.
- **Currency and Completeness.** The U.S. Census Bureau has sought to link the subnational population data to the most current and complete shapefiles of those subnational areas obtainable. The ideal shapefile is that of subnational areas that correspond exactly to those at a country's most recent census.

- **Retrofitting.** Some subnational areas in a country shapefile may differ from those reported in the country's most recent census. Usually this is because the shapefile is of a vintage that predates the formation of new subnational areas out of older areas as reported in the census. Records are added to the spreadsheet with matching population data for the "older" polygons. Such records usually are aggregates of the "newer" subnational areas that fit the "old."
- **Shapefile Sources.** The U.S. Census Bureau has drawn upon various sources for the shapefiles, having modified or updated the attribute data for some of them. Such sources include these:
 - United Nations *Second Administrative Level Boundary (SALB)* files (http://www.who.int/whosis/database/gis/salb/salb_home.htm)
 - United Nations Environment Programme/Global and Regional Integrated Data Centres (UNEP/GRID, the originals from the early 1990s)
 - Consultative Group on International Agricultural Research's Consortium for Spatial Information/Centro Internacional de la Papa (CGIAR/CIP, <http://research.cip.cgiar.org/gis/index.php>)
 - University of California, Berkeley, *Global Administrative Division Areas (Vs. 6, <http://biogeo.berkeley.edu/qadm>)*
 - Map Maker Trust (Scotland), *Map Library* (<http://www.maplibrary.org>)
 - U.S. Department of State Office of the Geographer and Global Issues
 - National statistical offices (e.g., Malawi, South Africa)
 - A combination of any of these
- **Disclaimer.** Many of the shapefiles come from open sources, including those that are readily downloadable from specific websites. The U.S. Census Bureau maintains these files in the Geographic Coordinate System GCS WSG 1984. Except in rare instances, the U.S. Census Bureau has made no attempt to modify digital boundaries, nor can it ensure their positional accuracy.

5. Spreadsheets

- **Comprehensive Listing of Population Data at all Administrative Levels.** The subnational 5-year age/sex group population estimates/projections for the country and its subnational areas are contained in a single stand-alone spreadsheet. This includes records that provide the direct linkage to the country shapefile polygons (joined through the "GEO_MATCH" field), some being the aggregates of multiple subnational areas.
- **Individual Worksheets for the Individual Years.** Each spreadsheet contains individual worksheets for each year between 2000 and 2015. The data in each worksheet are joined to the corresponding polygons in the shapefile through the "GEO_MATCH" field.
- **Data Dictionary.** This is provided in a separate worksheet. (It also is appended at the end of this document.)

6. Metadata

- The metadata files were created from the ArcGIS. They comply with the **U.S. Federal Geographic Data Committee** standards.

- The metadata files have information on the following: **citations** of the census data, the shapefiles, and the U.S. Census Bureau *International Data Base*; the shapefiles; the **administrative divisions** within a country and any supplemental background pertaining to that country; **spatial characteristics** of the shapefile; and **contact information**.

7. Standards

As mentioned in the previous pages, the data that the U.S. Census Bureau is providing in the HIV Spatial Data Repository Repository adhere to the following standards:

- **Geographic Area Names.** Spellings from the U.S. Board on Geographic Names (BGN), available on the National Geospatial-Intelligence Agency's *GEOnet Names Server* website <http://gnswww.nga.mil/geonames/GNS/index.jsp>, are used for the subnational areas. When a BGN spelling is unavailable, the spelling from the country's census is used.
- **Geographic Area Boundaries.** National boundaries adhere to those recognized by the Office of the Geographer and Global Issues of the U.S. Department of State.
- **Metadata.** The metadata files generated from their respective shapefiles adhere to the U.S. Federal Geographic Data Committee (FGDC) standards.

8. Updates, Upgrades, Additions

The U.S. Census Bureau will upload new, expanded, or revised data to the HIV Spatial Data Repository Repository on a flow basis. Such data may consist of any of the following:

- New 5-year age/sex group estimates/projections for a country based on the availability of new census results
- New 5-year age/sex group estimates/projections for countries following a release of updated national projections in the U.S. Census Bureau *International Data Base*
- A country shapefile with new subnational boundaries or with boundaries of the next lower level administrative division, which will supersede the older shapefile
- Subnational population data and shapefiles of any countries added to those receiving PEPFAR sponsorships

DATA DICTIONARY

The following are descriptions of the fields in both the country shapefiles and spreadsheets.

<u>FIELD NAME</u>	<u>DESCRIPTION</u>
GEO_MATCH	The field that provides the matching link between the spreadsheets and shapefiles – unique for each record. This is a concatenation of administrative area names at all levels separated by an underscore (“_”), beginning with the country name and ending with the lowest administrative level of the administrative area. A polygon that consists of an aggregate of 2 or more administrative areas has within its name the names of those areas separated by a plus sign (“+”).
GEO_AGG	The label for the areas that are components of an aggregated area in the shapefile, which is the same as that in the GEO_MATCH field for the aggregate record.
GEO_LABEL	The listing of the names of all administrative areas at all levels in one column. Names are in the order of the lowest level nested within the next higher level, i.e., the ADM3s within each ADM2, the ADM2s within each ADM1, and the ADM1s within the COUNTRY.
COUNTRY	The country name
ADM1_NAME	The first-level administrative area (e.g., province or state)
ADM2_NAME	The second-level administrative area (e.g., district or county)
ADM3_NAME	The third-level administrative area (e.g., township or arrondissement)
ADM_LEVEL	<p>“0” -- country level “1” – ADM1 level “2” -- ADM2 level “3” -- ADM3 level</p> <p>The code for a polygon representing multiple administrative areas includes the number of those areas in parentheses. The code for a polygon at the second administrative level representing 3 areas is “2(3).”</p> <p>The code for a polygon representing areas at different administrative levels is shown with the levels separated by a forward slash (“/”). For a polygon with 1 second-level area and 2 third-level areas, the code is “2/3(2).”</p>
COMMENT	Any notes about an area, which may include former names or alternate names or spellings
FIPS	The U.S. Government Federal Information Processing System (FIPS) code (for country and first-level administrative division records only)
STAT_CODE	An official code of an administrative area that is provided by the country’s statistical agency (where available), usually in the national census
SALB_CODE	Administrative area codes provided in the World Health Organization’s <i>Second Administrative Level Boundary</i> shapefiles (where available)

OTHER_CODE	Any other publicly available administrative area code in use by the PEPFAR community
BTOTL_YYYY	The <i>total</i> population estimate (both sexes) for the year “YYYY”
B0004_YYYY	The population estimate of <i>all people</i> (both sexes) between ages 0 and 4 for the year “YYYY”
B0509_YYYY	The population estimate of <i>all people</i> (both sexes) between ages 5 and 9 for the year “YYYY”
...	...
B7579_YYYY	The population estimate of <i>all people</i> (both sexes) between ages 75 and 79 for the year “YYYY”
B80PL_YYYY	The population estimate of <i>all people</i> (both sexes) at the ages of 80 and over for the year “YYYY”
MTOTL_YYYY	The total <i>male</i> population estimate for the year “YYYY”
M0004_YYYY	The population estimate of <i>males</i> between ages 0 and 4 for the year “YYYY”
M0509_YYYY	The population estimate of <i>males</i> between ages 5 and 9 for the year “YYYY”
...	...
M7579_YYYY	The population estimate of <i>males</i> between ages 75 and 79 for the year “YYYY”
M80PL_YYYY	The population estimate of <i>males</i> at the ages of 80 and over for the year “YYYY”
FTOTL_YYYY	The total <i>female</i> population estimate for the year “YYYY”
F0004_YYYY	The population estimate of <i>females</i> between ages 0 and 4 for the year “YYYY”
F0509_YYYY	The population estimate of <i>females</i> between ages 5 and 9 for the year “YYYY”
...	...
F7579_YYYY	The population estimate of <i>females</i> between ages 75 and 79 for the year “YYYY”
F80PL_YYYY	The population estimate of <i>females</i> at the ages of 80 and over for the year “YYYY”