EPPSA SECONDARY DATA MANUAL

VERSION 2 – UPDATED JUNE 2021

OVERVIEW

The EPPSA project takes a collaborative, interdisciplinary approach to research multilevel determinants of energy poverty, the effects of energy poverty on energy poor populations, the impacts of fuel burning on air quality and natural resources, factors in energy transitions in sub-Saharan Africa, and functionality of health systems and other public infrastructure in societies with limited electricity infrastructure. In addition to our primary data collection efforts, we maintain a comprehensive secondary database to answer research questions about energy poverty in sub-Saharan Africa. Our team gathered and cataloged population representative survey data, geospatial infrastructure data, and remotely sensed datasets to test hypotheses related to meso- and macro-level determinants of energy poverty and impacts of shocks or natural experiments, and provide contextual data regarding how longer-term trends in population and environment dynamics are influencing energy poverty related outcomes. These datasets were produced and are owned by entities not associated with our project, so our aim is to produce a database that connects researchers to data that may be outside of their discipline.

This document is to serve as a source of information about the different secondary datasets that have been gathered by our team from various sources. The datasets are organized into different categories and by data format. Over time, this manual will be updated with the most current version. In this document, you can find: what data our team currently has access to, for which countries and years, and how to access the datasets.

Special thanks to Dr. Yu Wu, Ryan McCord, Cyrus Sinai, Daniel Han, Madeline Chandler, Dr. Michael Emch, and the members of the Spatial Health Research Group at UNC for their efforts in building this secondary database. If you would like to access any of the data described here or have suggestions of datasets to be added, email Kate Brandt (kebrandt@live.unc.edu) for assistance.

Summary of major updates:

- Additions: High Frequency Phone Survey 2020-21 for Malawi, SPA DHS data, standardized national indicators (air pollution, disease burdens), ‘rnoaa’ package description
- Addition of descriptions on how to download datasets

DIRECTORY

I. HOUSEHOLD SURVEYS

INTRODUCTION

Household surveys are used to evaluate population-level characteristics for countries. Each survey described in this section was designed and administered with a specific purpose. For example, the Demographic and Health Surveys (DHS) are administered regularly with an emphasis on understanding population-level health characteristics in addition to gathering information about poverty and quality of life characteristics; the Living Standards Measurement Study (LSMS) focuses on household economics and assets. Each survey may be more appropriate for a given secondary-data study based on the depth of the modules, survey design (i.e., cross sectional or panel surveys), and whether or not the survey includes geospatial location information. **Table 1** summarizes the survey datasets that are currently available to our team, available by year and country. Note: though the emphasis of the secondary-data gathering project has been on the EPPSA study countries, our team currently has access to surveys of other countries, listed in **Table 1**.

The datasets currently available to our team include:

- Demographic and Health Surveys (**DHS**)
- Living Standards Measurement Study (**LSMS**)
- Multiple Indicator Cluster Surveys (**MICS**)
- High Frequency Phone Survey (**HFPS**)
- Malawi Longitudinal Study of Families and Health (**MLSFH**)
- Multi-Tier Framework (**MTF**)

There is also a spreadsheet that details energy-poverty relevant questions asked in the DHS, LSMS, and MICS in ***Appendix A***.

Table 1. Summary household surveys currently available to our team for EPPSA study countries

Year	Malawi	Zambia	Zimbabwe	DR Congo	Ghana	Gabon
1992	DHS	DHS				
1993					DHS*	
1994			DHS			
1995						
1996		DHS				
1997						
1998	MLSFH				DHS*	
1999		MICS	DHS*			
2000	DHS*					DHS
2001	MLSFH	DHS*				
2002						
2003					DHS*	
2004	DHS* LSMS MLSFH					
2005			DHS*			
2006	MICS* MLSFH				MICS	
2007		DHS*		DHS *		
2008	MLSFH				DHS*	
2009			MICS*			
2010	DHS* LSMS* MLSFH		DHS*			
2011					MICS*	
2012	DHS* MLSFH					DHS*
2013	DHS* LSMS* MICS*	DHS*		DHS *		
2014			MICS*		DHS*	
2015	DHS*		DHS*			
2016	LSMS*				DHS*	
2017	DHS*	MTF		MICS*	DHS*;MICS*	
2018		DHS*				
2019	LSMS*		MICS*		DHS*	
2020	HFPS*					

Note: Many surveys begin in one year with data collection continuing through the following year. Only the year in which a wave of data collection began is included in this table.

*Indicates GPS coordinates are available.

DEMOGRAPHIC AND HEALTH SURVEYS (DHS)

Overview

The DHS are nationally-representative surveys that provide data for a wide range of monitoring and impact evaluation indicators in the areas of population, health, and nutrition. The standard DHS collects information about household assets, household member characteristics, maternal and child health, attitudes towards family planning and women's empowerment, and knowledge and behaviors regarding malaria and HIV/AIDS. Surveys may include different questions and modules depending on the country and survey year. Some DHS waves collect biomarkers, such as blood spots.

Owner	USAID
Website	https://dhsprogram.com/data/available-datasets.cfm
Unit of Analysis	Household, Individual
GPS Coordinates	Household clusters
Sample Size	5,000-30,000 households
Coverage	Nationally representative

Geographic Data

Household clusters are georeferenced with offset coordinates to maintain confidentiality. GPS latitude and longitude coordinates are offset by up to 2 kilometers in urban areas and 5 kilometers in rural areas (with 1% of rural community coordinates offset by up to 10 kilometers). The displacement is restricted so that the offset coordinates are still within the country and DHS survey region.

Relevance to Energy Poverty Research

The DHS provide detailed information on health outcomes, including maternal and child health, infant mortality, and respiratory health outcomes– all of which are of interest to research on the health impacts of fuel burning. The household surveys contain information about a household's fuel choices and electricity access, but these questions are not as detailed as other nationally-representative surveys (like LSMS).

Other Notes

One advantage of DHS data are the geographic coverage (over 90 countries surveyed) and the frequency with which data are collected (roughly every 5 years). This frequency and coverage can make cross-country comparison feasible, but sometimes questions are changed by country to fit the population's needs. This can make data processing for a cross-country or even year-to-year comparison tedious. The IPUMS (University of Minnesota) makes analyzing DHS data from multiple countries and survey waves easier by maintaining a database of consistently coded variables: https://www.idhsdata.org/idhs/

How to Access

Register an account on the DHS site (https://dhsprogram.com/data/new-user-registration.cfm). With a registered account, you may download any of the publicly available datasets.

LIVING STANDARDS MEASUREMENT STUDY (LSMS)

Overview

The Living Standards Measurement Study (LSMS) is the World Bank's flagship household survey program focused on collecting and analyzing data on household and individual wellbeing to better inform development policies. The LSMS team is housed in the Data Production and Methods Unit of the World Bank's Development Data Group.

Owner	World Bank
Website	https://www.worldbank.org/en/programs/lsms
Unit of Analysis	Household, Individual, Community
GPS Coordinates	Enumeration Areas (clusters)
Sample Size	1,500-18,000 households; 100-300 Enumeration Areas (communities)
Coverage	Nationally representative

Geographic Data

Household Enumeration Areas (EA) are georeferenced with offset coordinates to maintain confidentiality. GPS latitude and longitude coordinates are offset by up to 2 kilometers in urban areas and 5 kilometers in rural areas (with 1% of rural community coordinates offset by up to 10 kilometers). The displacement is restricted so that the offset coordinates are still within the country and DHS survey region. Geospatial variables included in the LSMS datasets have been produced using the unmodified GPS coordinates for an EA; these variables include distances to roads and markets, climate information, and information about land use and land cover.

Relevance to Energy Poverty Research

The LSMS includes a module on household energy use which asks detailed questions about household fuel choices, stacking, gathering, electricity access, and stove types, and cooking practices. In addition to household energy data, the LSMS includes detailed information about employment and labor of the household members, education attainment, household assets and business enterprises, and some questions about health.

Other Notes

For some countries, a subset of households surveyed in the LSMS are followed as part of a multi-year panel study. These panel data can be extremely useful for assessing change over time. The Malawi Integrated Household Survey has a short version with two waves (2010-2013; 4,000 households) and a long version with four waves of data (2010-2013-2016-2019; 3,000 households).

How to Access

Search The World Bank Microdata Library: https://microdata.worldbank.org/index.php/catalog. You may filter for study (LSMS) and country. To download a dataset, click on the "Get Microdata" tab. Register an account or login to your existing account to download.

MULTIPLE INDICATOR CLUSTER SURVEY (MICS)

Overview

The Multiple Indicator Cluster Surveys (MICS) is an international household survey program developed by UNICEF to collect statistically sound, internationally comparable estimates of wellbeing indicators for children, men, and women. The MICS do not typically collect biomarkers but do collect anthropometric data for all children under five years old.

Owner	UNICEF
Website	http://mics.unicef.org
Unit of Analysis	Household, Individual
GPS Coordinates	Household cluster level
Sample Size	4,000-8,000 households
Coverage	Regional or sub-population representative sample

Geographic Data

From the MICS site: In the sixth round of MICS, countries have the option to collect GIS data on the location of survey clusters where interviews take place. Even if countries do not collect such data in a MICS survey, such data are usually available from the majority of national statistical offices which usually have digitized maps of cluster locations through their census cartography. The MICS Programme therefore advises researchers interested in spatial analysis to contact the individual statistical offices or other implementing agencies with requests. Contact details are typically in final reports and with the final datasets. One needs the “key” that matches the cluster numbers in the datasets with the enumeration areas in the Census maps. Additionally, one must be granted access to Census maps and, for any map presentation, must incorporate a random offset of the cluster location.

Relevance to Energy Poverty Research

The MICS include questions about fuelwood collection, cooking practices, stove type, household fuel choice, household labor and assets which are all relevant to understanding household fuel use. The MICS data are a rich source of data for analyzing progress towards the United Nation’s Sustainable Development Goals, so data collection on health and education and priorities for the surveys.

How to Access

Search for datasets by country and year on the MICS site: https://mics.unicef.org/surveys. Register for an account to download datasets.

HIGH FREQUENCY PHONE SURVEYS (HFPS)

Overview

The HFPS was implemented in a few countries based on households in the LSMS surveys. In Malawi, nine rounds of phone surveys were implemented from 2020-2021 to ask households about their responses to the COVID-19 pandemic. In Zambia, one rapid phone survey was implemented in June 2020.

Owner	World Bank
Website	https://www.worldbank.org/en/programs/lsms/brief/lsms-launches-high-frequency-phone-surveys-on-covid-19
Unit of Analysis	Household, Individual
GPS Coordinates	Enumeration Areas (clusters)
Sample Size	1,500-2,000 households
Coverage	Nationally representative

Geographic Data

Enumeration Areas (EA) from LSMS surveys. Household EAs are georeferenced with offset coordinates to maintain confidentiality. GPS latitude and longitude coordinates are offset by up to 2 kilometers in urban areas and 5 kilometers in rural areas (with 1% of rural community coordinates offset by up to 10 kilometers). The displacement is restricted so that the offset coordinates are still within the country and DHS survey region. Geospatial variables included in the LSMS datasets have been produced using the unmodified GPS coordinates for an EA; these variables include distances to roads and markets, climate information, and information about land use and land cover.

Relevance to Energy Poverty Research

Households that were able to be reached by phone contact were asked about their experiences during the first year of the COVID-19 pandemic. Questions were asked about children’s access to education through virtual learning, access to basic services and health services, employment, agriculture, and shocks.

How to Access

Search The World Bank Microdata Library: https://microdata.worldbank.org/index.php/catalog. You may filter for study (HFPS) and country. To download a dataset, click on the “Get Microdata” tab. Register an account or login to your existing account to download.

MALAWI LONGITUDINAL STUDY OF FAMILIES AND HEALTH (MLSFH)

Overview

The Malawi Longitudinal Study of Families and Health (MLSFH) is one of very few long-standing publicly-available longitudinal cohort studies in a sub-Saharan African (SSA) context. It provides a rare record of more than a decade of demographic, socioeconomic and health conditions in one of the world's poorest countries. With 7 data collection rounds spanning 1998 to 2012 for up to 4,000 individuals, the MLSFH permits researchers to investigate the multiple influences that contribute to HIV risks in sexual partnerships, the variety of ways that people manage risk within and outside of marriage, the possible effects of HIV prevention policies and programs, and the mechanisms through which poor rural individuals, families, households, and communities cope with the impacts of high morbidity and mortality that are often—but not always—related to HIV/AIDS.

Owner	Population Studies Center and the University of Pennsylvania
Website	https://malawi.pop.upenn.edu/
Unit of Analysis	Individual
GPS Coordinates	None
Sample Size	Up to 4,000
Coverage	Regional (based in Rumphi, Mchinji, and Balaka)

Relevance to Energy Poverty Research

There are no questions directly related to household energy use and cooking. You may want to use the panel data to study the relationships between household dynamics and environmental change through use of ancillary climate and land use/land cover datasets.

How to Access

From the MLSFH site: https://malawi.pop.upenn.edu/malawi-data-mlsfh
A public-use version of the 1998-2010 MLSFH data without identifying individual or village information is currently processed for inclusion at the ICPSR at the University of Michigan. In the interim, until ICPSR has released the data, researchers interested in working with the MLSFH data should send a short project description and a signed copy of the **MLSFH data use agreement** to Hans-Peter Kohler (hpkohler@pop.upenn.edu), and a link with the MLSFH public-use data will be emailed.

MULTI-TIER FRAMEWORK SURVEY FOR MEASURING ENERGY ACCESS (MTF)

Overview

The Multi-Tier Framework (MTF) initiative launched in June 2015 by a sector of the World Bank group to understand electricity access across a spectrum of service levels. Data collection for the Zambia MTF occurred from September 2017 to March 2018. The survey's objective is to provide more nuanced data on energy access, including access to electricity and cooking solutions. The MTF approach goes beyond the traditional binary measurement of energy access to capture the multidimensional nature of energy access and the vast range of technologies and sources that can provide energy access, while accounting for the wide differences in user experience.

Owner	World Bank
Website	https://microdata.worldbank.org/index.php/catalog/3527
Unit of Analysis	Household
GPS Coordinates	Household cluster level
Sample Size	3,600
Coverage	Nationally representative

Geographic Data

Coordinates of enumeration areas (EAs) are available upon request. Email Kate (kebrandt@live.unc.edu) to get coordinate information.

Relevance to Energy Poverty Research

The dataset contains responses from households to questions on experiences concerning their electricity services and cooking practices, as well as questions on other basic socioeconomic factors, such as age, gender, education, expenditure, and health.

How to Access

Search The World Bank Microdata Library: https://microdata.worldbank.org/index.php/catalog. You may filter for study (MTF) and country; currently MTF data are available for Zambia, Nigeria, Nepal, Kenya, Honduras, and Niger. To download a dataset, click on the “Get Microdata” tab. Register an account or login to your existing account to download.

II. ECOLOGICAL AND CLIMATE DATA

INTRODUCTION

Ecological and climate data can be integrated with survey data or emissions data to identify large scale patterns, such as deforestation, changes in land use/land cover, and emissions exposure at a regional level. These data can also be used as community level covariates in models to assess a household's access to fuelwood which may help explain some behavior identified in a household survey dataset.

ELEVATION

SHUTTER RADAR TOPOGRAPHY MISSION (SRTM)

Land topography allows us to make maps of the features of the surface of the Earth. Topographic maps show the location, height, and shape of features like mountains and valleys, rivers, even the craters on volcanoes. Incorporation of altitude data into studies can help account for differences impacts of air pollution on health and assess differences in land use and land cover for communities living at different altitudes.

Owner	US Geological Survey		
Website	https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1-arc?qt-science_center_objects=0#qt-science_center_objects		
Data Format	TIFF		
Product Specifications			
Projection	Geographic	**Spatial Resolution**	1 arc-second (~30 meters)
Horizontal Datum	WGS84		3 arc-seconds (~90 meters)
Vertical Datum	EGM96	**Raster Size**	1 degree tiles
Vertical Units	Meters	**C-band Wavelength**	5.6 cm
Coverage	Global		

Details

This image was derived from the U.S. Geological Survey's GTOPO30 data set. GTOPO30 is a global digital elevation model (DEM) resulting from a collaborative effort led by the staff at the U.S. Geological Survey's EROS Data Center in Sioux Falls, South Dakota. Elevations in GTOPO30 are regularly spaced at 30-arc seconds (approximately 1 kilometer). GTOPO30 was developed to meet the needs of the geospatial data user community for regional and continental scale topographic data.

GTOPO30 is a global data set covering the full extent of latitude from 90 degrees south to 90 degrees north, and the full extent of longitude from 180 degrees west to 180 degrees east. The horizontal grid spacing is 30-arc seconds (0.008333 degrees), resulting in a DEM having dimensions of 21,600 rows and 43,200 columns. The horizontal coordinate system is decimal degrees of latitude and longitude referenced to WGS84. The vertical units represent elevation in meters above mean sea level. The elevation values range from -407 to 8,752 meters. In the DEM, ocean areas have been masked as "no data" and have been assigned a value of -9999. Lowland coastal areas have an elevation of at least 1 meter, so in the event that a user reassigns the ocean value from -9999 to 0 the land boundary portrayal will be maintained. Due to the nature of the raster structure of the DEM, small islands in the ocean less than approximately 1 square kilometer will not be represented.

The relief shading in this topographic map comes mostly from elevation data collected by space-based radars. A radar in space sends a pulse of radio waves toward Earth and measures the strength and length of time it takes a signal to bounce back. From this information, we can derive the height and shape of the features on the surface. GTOPO30 provides a best level of detail in global topographic data that are publicly available. GTOPO30 data are suitable for many regional and continental applications, such as climate modeling, continental-scale land cover mapping, extraction of drainage features for hydrologic modeling, and geometric and atmospheric correction of medium- and coarse-resolution satellite image data.

How to Access

Use USGS Earth Explorer (https://earthexplorer.usgs.gov/) to identify the geographic coverage and dataset for download (Digital Elevation > SRTM > SRTM 1 Arc-Second Global). Once you have chosen the dataset and area of interest, click the “Results” button to add data extracts to your cart. Register an account or login at https://ers.cr.usgs.gov/login to add data to your cart to download.

FIRES

VIIRS (S-NPP) I BAND 375 M ACTIVE FIRE PRODUCT NRT

Vector data that contain detected fire incidence from satellite acquisition. The vector data have a number of attributes including information about the satellite data collection, time and date of fire detection, and inferred hot spot type (0 = presumed vegetation fire, 1 = active volcano, 2 = other static land source, 3 = offshore detection (includes all detections over water)). Fire data can be used to understand large scale emissions patterns and changes in land use patterns.

Owner	National Aeronautics and Space Administration (NASA)		
Website	https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/v1-vnp14imgt		
Data Format	Vector (shapefile, KML, TXT, CSV, and JSON)		
Product Specifications			
Projection	WGS84	**Spatial Resolution**	375 m
Coverage	Global	**Time Interval**	Daily*
Years	2012-present		*Near Real Time

Details

Near real-time (NRT) Suomi National Polar-orbiting Partnership (Suomi NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) Active Fire detection product is based on that instrument's 375 m nominal resolution data. Compared to other coarser resolution (≥1km) satellite fire detection products, the improved 375 m data provide greater response over fires of relatively small areas, as well as improved mapping of large fire perimeters. Consequently, the data are well suited for use in support of fire management (e.g., near real-time alert systems), as well as other science applications requiring improved fire mapping fidelity. The 375 m product complements the baseline Suomi NPP/VIIRS 750 m active fire detection and characterization data, which was originally designed to provide continuity to the existing 1 km Earth Observing System Moderate Resolution Imaging Spectroradiometer (EOS/MODIS) active fire data record. Due to frequent data saturation issues, the current 375 m fire product provides detection information only with no sub-pixel fire characterization.

How to Access

To download vector data of fires in the past 24 hours to 7 days, use this link: https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data

To download data for fires before the previous 7 days, use this link to submit an archive download request: https://firms.modaps.eosdis.nasa.gov/download/

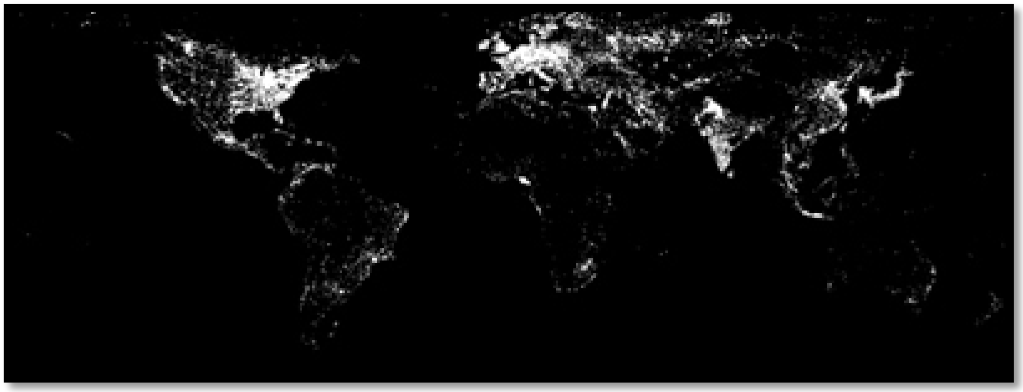
NIGHTTIME LIGHTS

DMSP-OLS NIGHTTIME LIGHTS VERSION 4

The nighttime lights product known as avg_lights_x_pct is derived from the average visible band digital number (DN) of cloud-free light detections multiplied by the percent frequency of light detection. The inclusion of the percent frequency of detection term normalizes the resulting digital values for variations in the persistence of lighting. For instance, the value for a light only detected half the time is discounted by 50%. Note that this product contains detections from fires and a variable amount of background noise. This is the product used to infer gas flaring volumes from the nighttime lights.

Nighttime lights have been used as a proxy for urban areas in studies where no reliable population data are available.

Owner	National Centers for Environmental Information (NOAA)		
Website	https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html#AVSLCFC		
Data Format	GeoTIFF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	30 arc-sec
Years	1992-2013	**Time Interval**	Year



2003 Nighttime Lights Composite

How to Access

Download global coverage GeoTIFF files for years of interest directly from https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html#AXP

NORMALIZED VEGETATION INDEX (NDVI)

MODIS/TERRA VEGETATION INDICES MONTHLY L3 GLOBAL

This is a derived product from MODIS imagery. Vegetation indices are used for global monitoring of vegetation conditions and are used in products displaying land cover and land cover changes. These data may be used as input for modeling global biogeochemical and hydrologic processes as well as global and regional climate. Additional applications include characterizing land surface biophysical properties and processes, such as primary production and land cover conversion.

Provided along with the vegetation layers and the two quality assurance (QA) layers are reflectance bands 1 (red), 2 (near-infrared), 3 (blue), and 7 (mid-infrared), as well as three observation layers.

Owner	National Aeronautics and Space Administration (NASA)		
Website	https://lpdaac.usgs.gov/products/mod13a3v006/		
Data Format	BIL, DTED, GeoTIFF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	1 km
Years	February 2000-present	**Time Interval**	Month

Details

The Terra Moderate Resolution Imaging Spectroradiometer (MODIS) Vegetation Indices (MOD13A3) Version 6 data are provided monthly at 1 kilometer (km) spatial resolution as a gridded Level 3 product in the sinusoidal projection. In generating this monthly product, the algorithm ingests all the MOD13A2 products that overlap the month and employs a weighted temporal average.

The MODIS Normalized Difference Vegetation Index (NDVI) complements NOAA's Advanced Very High Resolution Radiometer (AVHRR) NDVI products and provides continuity for time series historical applications. MODIS also includes an Enhanced Vegetation Index (EVI) that minimizes canopy background variations and maintains sensitivity over dense vegetation conditions. The EVI uses the blue band to remove residual atmosphere contamination caused by smoke and sub-pixel thin clouds. The MODIS NDVI and EVI products are computed from surface reflectance corrected for molecular scattering, ozone absorption, and aerosols.

How to Access

Use USGS Earth Explorer (https://earthexplorer.usgs.gov/) to identify the geographic coverage and dataset for download (NASA LPDAAC Collections > MODIS Vegetation Indices - V6 > MODIS MOD13A3 V6). Once you have chosen the dataset and area of interest, click the “Results” button to add data extracts to your cart. Register an account or login at https://ers.cr.usgs.gov/login to add data to your cart to download.

SOILS

HARMONISED WORLD SOIL DATABASE – WISE 30SEC SOIL DATA SET

The GIS dataset was created using the soil map unit delineations of the broad scale Harmonised World Soil Database, version 1.21, with minor corrections, overlaid by a climate zones map (Köppen-Geiger) as co-variate, and soil property estimates derived from analyses of the ISRIC-WISE soil profile database for the respective mapped 'soil/climate' combinations.

The dataset considers 20 soil properties that are commonly required for global agro-ecological zoning, land evaluation, crop growth simulation, modelling of soil gaseous emissions, and analyses of global environmental change. It presents 'best' estimates for: organic carbon content, total nitrogen, C/N ratio, pH(H2O), CECsoil, CECclay, effective CEC, total exchangeable bases (TEB), base saturation, aluminium saturation, calcium carbonate content, gypsum content, exchangeable sodium percentage (ESP), electrical conductivity, particle size distribution (content of sand, silt and clay), proportion of coarse fragments (less than 2 mm), bulk density, and available water capacity (-33 to -1500 kPa); also the dominant soil drainage class.

Owner	ISRIC – World Soil Information
Website	https://www.isric.org/explore/wise-databases
Data Format	GeoTIFF
Product Specifications	
Projection	Coordinate Reference System - EPSG:4326
Coverage	Global
Spatial Resolution	30 arc-seconds

Details

Soil property estimates are presented for fixed depth intervals of 20 cm up to a depth of 100 cm, respectively of 50 cm between 100 cm to 200 cm (or less when appropriate) for so-called 'synthetic' profiles' (as defined by their 'soil/climate' class). The respective soil property estimates were derived from statistical analyses of data for some 21,000 soil profiles managed in a working copy of the ISRIC-WISE database; this was done using an elaborate scheme of taxonomy-based transfer rules complemented with expert-rules that consider the 'in-pedon' consistency of the predictions. The type of rules used was flagged to provide an indication of the possible confidence (i.e. lineage) in the derived data.

Best estimates for each attribute are given as means and standard deviations (STD), as calculated for the sample populations that remained upon application of a robust data outlier detection scheme. Results of the analyses can be linked to the spatial data through the unique map unit (grid cell) identifier, which is a combination of the soil unit and climate class code. Most map units are comprised of up to ten different components; each of these with their own range of derived soil properties and associated statistical uncertainties.

Estimates of global soil organic carbon (SOC) stocks to 200 cm are presented in the technical documentation as an example of possible application.

How to Access

Search the ISRIC Data Hub: “HWSD” https://data.isric.org/geonetwork/srv/eng/catalog.search#/home

FOREST COVER

GLOBAL FOREST CHANGE 2000-2019 V 1.7

This global dataset is divided into 10x10 degree tiles, consisting of seven files per tile. All files contain unsigned 8-bit values and have a spatial resolution of 1 arc-second per pixel, or approximately 30 meters per pixel at the equator. The dataset includes yearly information about tree canopy cover (from the year 2000 only), forest cover gain (cumulative 2000-2012), and forest cover loss (yearly 2001-2019). This is a derived product based on Landsat 8 data. This data has been incorporated into household survey analysis to understand household access to forest resources. It can be used to evaluate forest loss overtime, and forest gain in some time periods.

Owner	University of Maryland		
Website	https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html		
Data Format	TIFF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	1 arc-second (~30 meters) (approximately 30m at the equator)
Raster Size	10x10 degree tiles		
Years	2000-2019	**Time Interval**	Year

Details

Use the following credit when these data are cited:

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. “High-Resolution Global Maps of 21st-Century Forest Cover Change.” Science 342 (15 November): 850–53. Data available on-line from: https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html.

How to Access

Download directly from https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html. Point and click to identify 10x10 degree tiles of interest.

PRECIPITATION

CRU TS4.04

The Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate are monthly gridded fields based on monthly observational data calculated from daily or sub-daily data by National Meteorological Services and other external agents. All CRU TS output files are actual values - NOT anomalies. The CRU TS4.04 variables are **cloud cover, diurnal temperature range, frost day frequency, potential evapotranspiration (PET), precipitation, daily mean temperature, monthly average daily maximum and minimum temperature, and vapour pressure** for the period January 1901 - December 2019.

Owner	Center for Environmental Data Analysis (CEDA)		
Website	https://help.ceda.ac.uk/article/4472-cru-data-user-guide		
Data Format	ASCII or NetCDF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	0.5x0.5 degrees
Years	1901-2019	**Time Interval**	Month

Details

These data were produced by the Climatic Research Unit at the UK National Centre for Atmospheric Sciences and passed to the Centre for Environmental Data Analysis (CEDA).

Use the following credit when these data are cited:

University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D.; Osborn, T. (2020): CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2019). Centre for Environmental Data Analysis, *date of citation*. https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9

Additional information can be found on the CRU homepage: https://crudata.uea.ac.uk/cru/data/hrg/

How to Access

You may download CRU files through the web interface, OPeNDAP service, or file transfer protocol, depending on what formats and datasets you would like to download. Instructions can be found here: https://help.ceda.ac.uk/article/4472-cru-data-user-guide#Downloading%20the%20CRU%20data

TRMM/GPM 3B42 AND 3B43

The Tropical Rainfall Measuring Mission (TRMM) is a research satellite used by NASA to measure surface precipitation. The Global Precipitation Measurement Mission (GPM) uses satellites including TRMM to measure and produce estimates of precipitation. The TRMM/GPM 3B products are estimates of rainfall based on combined radar/radiometer data. The GPM Combined Radar-Radiometer Algorithm performs two basic functions: first, it provides, in principle, the most accurate, high resolution estimates of surface rainfall rate and precipitation vertical distributions that can be achieved from a spaceborne platform, and it is therefore valuable for applications where information regarding instantaneous storm structure are vital. Second, a global, representative collection of combined algorithm estimates will yield a single common reference dataset that can be used to “cross-calibrate” rain rate estimates from all of the passive microwave radiometers in the GPM constellation. The cross-calibration of radiometer estimates is crucial for developing a consistent, high time-resolution precipitation record for climate science and prediction model validation applications.

<table>
<tr><td><b>Owner</b></td><td colspan="3">Global Precipitation Measurement Mission (NASA)</td></tr>
<tr><td><b>Website</b></td><td colspan="3">https://gpm.nasa.gov/data/directory</td></tr>
<tr><td><b>Data Format</b></td><td colspan="3">HDF</td></tr>
<tr><td colspan="4"><b><i>Product Specifications</i></b></td></tr>
<tr><td><b>Coverage</b></td><td>Global</td><td><b>Spatial Resolution</b></td><td>0.25x0.25 degree</td></tr>
<tr><td><b>Years</b></td><td>Dec 1997 – Dec 2019</td><td><b>Time Interval</b></td><td>3 hours (3B42), month (3B43)</td></tr>
</table>

Details

There are differences between 3B42 and 3B43:

TRMM_3B42: TRMM (TMPA) Rainfall Estimate L3 3 hour 0.25 degree x 0.25 degree V7: This dataset is the output from the TMPA (TRMM Multi-satellite Precipitation Analysis) Algorithm, and provides precipitation estimates in the TRMM regions that have the (nearly-zero) bias of the ”TRMM Combined Instrument” precipitation estimate and the dense sampling of high-quality microwave data with fill-in using microwave-calibrated infrared estimates. The granule temporal coverage is 3 hours.
https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_7/summary

TRMM_3B43: TRMM (TMPA/3B43) Rainfall Estimate L3 1 month 0.25 degree x 0.25 degree V7: The 3B43 dataset is the monthly version of the 3B42 dataset.
This product is created using TRMM-adjusted merged microwave-infrared precipitation rate (in mm/hr) and root-mean-square (RMS) precipitation-error estimates.
It provides a "best" precipitation estimate in a latitude band covering 50o N to 50o S, an expansion of the TRMM region, from all global data sources, namely high-quality microwave data, infrared data, and analyses of rain gauges. The granule size is one month.
https://disc.gsfc.nasa.gov/datasets/TRMM_3B43_7/summary

How to Access

You must first register an account: https://registration.pps.eosdis.nasa.gov/registration/. Then, follow the instructions to download data either using FTPS or HTTPS: https://gpm.nasa.gov/data/directory

TEMPERATURE

CRU TS4.04

The Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high resolution gridded data of month-by-month variation in climate are monthly gridded fields based on monthly observational data calculated from daily or sub-daily data by National Meteorological Services and other external agents. All CRU TS output files are actual values - NOT anomalies. The CRU TS4.04 variables are **cloud cover, diurnal temperature range, frost day frequency, potential evapotranspiration (PET), precipitation, daily mean temperature, monthly average daily maximum and minimum temperature, and vapour pressure** for the period January 1901 - December 2019.

Owner	Center for Environmental Data Analysis (CEDA)		
Website	https://help.ceda.ac.uk/article/4472-cru-data-user-guide		
Data Format	ASCII or NetCDF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	0.5x0.5 degrees
Years	1901-2019	**Time Interval**	Month

Details

These data were produced by the Climatic Research Unit at the UK National Centre for Atmospheric Sciences and passed to the Centre for Environmental Data Analysis (CEDA).

Use the following credit when these data are cited:

University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D.; Osborn, T. (2020): CRU TS4.04: Climatic Research Unit (CRU) Time-Series (TS) version 4.04 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2019). Centre for Environmental Data Analysis, *date of citation*. https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9

Additional information can be found on the CRU homepage: https://crudata.uea.ac.uk/cru/data/hrg/

How to Access

You may download CRU files through the web interface, OPeNDAP service, or file transfer protocol, depending on what formats and datasets you would like to download. Instructions can be found here: https://help.ceda.ac.uk/article/4472-cru-data-user-guide#Downloading%20the%20CRU%20data

RADIATION

CERES-EBAF

The Clouds and Earth's Radiant Energy Systems (CERES) Energy Balanced and Filled (EBAF) product provides 1-degree regional, zonal and global monthly mean Top-of-Atmosphere (TOA) and surface (SFC) longwave (LW), shortwave (SW), and net (NET) fluxes under clear and all-sky conditions. EBAF is used for climate model evaluation, estimating the Earth's global mean energy budget, and to infer meridional heat transport.

Owner	Clouds and Earth's Radiant Energy Systems (CERES, NASA)		
Website	https://ceres.larc.nasa.gov/data/#energy-balanced-and-filled-ebaf		
Data Format	NC		
Product Specifications			
Coverage	Global	**Spatial Resolution**	1 degree grid
Years	Mar 2000 – Mar 2020	**Time Interval**	Month

Visit this page for an overview of the strengths and weaknesses of this dataset.

Loeb, Norman & National Center for Atmospheric Research Staff (Eds). Last modified 11 Jul 2018. **"The Climate Data Guide: CERES EBAF: Clouds and Earth's Radiant Energy Systems (CERES) Energy Balanced and Filled (EBAF)."** Retrieved from https://climatedataguide.ucar.edu/climate-data/ceres-ebaf-clouds-and-earths-radiant-energy-systems-ceres-energy-balanced-and-filled.

How to Access

From https://ceres.larc.nasa.gov/data/#energy-balanced-and-filled-ebaf, you may select which dataset(s) you are interested in downloading, toggle the map or enter coordinates to select geographic coverage, and select time frame. You will need to register an account to download the data.

R PACKAGE TO EASILY DOWNLOAD DATA FROM NOAA

If you are savvy in R, the '**rnoaa**' package is an easy way to download datasets from the National Oceanic and Atmospheric Administration. The **rnoaa** package allows users to download data on rainfall, severe weather, tornadoes, sea ice, and more.

Read more about the package and its capabilities at https://docs.ropensci.org/rnoaa/

III. POPULATION DATA

INTRODUCTION

Survey data from Section I provides for crucial information about population health, human behavior, and household economics. The population data described here are gridded population data products that model human population density over the Earth's surface. Where census data has not been recently or regularly collected, these products model population estimates based on other geographic data including land use/land cover, vegetation, and infrastructure. These data can be used to assess population change on a regional level or population density to characterize community level dynamics.

Additionally, national-level trend indicators of the burden of air pollution are useful for comparing different country trends and providing context to a study area. The WHO Global Health Observatory data dashboard has many indicators of health which are grouped by theme that can be downloaded for one or many countries over time.

Table 2. Description of gridded population products (modified from https://www.popgrid.org/data-docs-table1)

Owner	Concept	Spatial Resolution	Years
Unmodeled Population Grids			
Gridded Population of the World (GPW), version 4			
Center for International Earth Science Information Network (CIESIN)	Nighttime population (population counted at place of domicile)	30 arc-seconds (1 km)	2000, 2005, 2010, 2015, 2020
Lightly Modeled Population Grids			
Global Human Settlement Layer – Population (GHS-POP)			
Joint Research Centre and CIESIN	Nighttime population (population counted at place of domicile)	9 arc-seconds (~250 m), 30 arc-seconds (~1 km), WGS84	1975, 1990, 2000, 2015
Global Rural Urban Mapping Project (GRUMP)			
CIESIN; International Food Policy Research Institute; The World Bank; Centro Internacional de Agricultural Tropical	Nighttime population (population counted at place of domicile)	**30 arc-seconds (1 km)**	1990, 1995, 2000
Highly Modeled Population Grids			
LandScan Global Population database			
Oak Ridge National Laboratory	Day time (ambient) population	30 arc-seconds (1 km)-	annual releases 2000 - 2016 (current version)
WorldPop			
WorldPop	High spatial resolution, temporally-explicit data on human population and demographic distributions	3-arc seconds (100 meter)	2000-2020 globally and country-specific years

* Dasymetric mapping approaches rely on ancillary data to spatially disaggregate census counts from administrative/census units. The simplest approach is binary dasymetric mapping, which uses one other data layer (such as satellite-derived built-up areas or urban extents) to move populations from census units (which are sometimes large) to areas identified as settlements.

GRIDDED POPULATION DATA

WORLDPOP

WorldPop develops peer-reviewed research and methods for the construction of open and high-resolution geospatial data on population distributions, demographics and dynamics, with a focus on low and middle income countries. Datasets developed at WorldPop include information about estimated population counts, age sex structures, births, pregnancies, internal migration, and other development indicators.

Owner	WorldPop		
Website	https://www.worldpop.org/		
Data Format	TIFF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	100 m
Years	2000-2020	**Time Interval**	Year

Summary of Variables
Variables that WorldPop produces datasets for: estimated population counts, age sex structures, births, pregnancies, internal migration, and other development indicators

How to Access
Visit https://www.worldpop.org/ and download data directly from the “Data” tab. You may download data based on variable of interest or country.

ACCESSIBILITY TO CITIES

Quantify and validate global accessibility to high-density urban centres at a resolution of 1×1 kilometre for 2015, as measured by travel time. The last global mapping effort to measure accessibility was for the year 2000, a time that predates both substantial investment and expansion of transportation infrastructure and an extraordinary improvement in the data quantity and quality of accessibility measures.

Owner	The Malaria Atlas Project		
Website	https://malariaatlas.org/research-project/accessibility_to_cities/		
Data Format	TIFF		
Product Specifications			
Coverage	Global	**Spatial Resolution**	1 km
Years	2015	**Time Interval**	Year

How to Access

Visit https://malariaatlas.org/research-project/accessibility-to-cities/ and download data directly from the bottom of the page.

NATIONAL TREND INDICATORS

WHO GLOBAL HEALTH OBSERVATORY DATA DASHBOARD

The WHO GHO data dashboard has information that ranges from burden of indoor air pollution, to infectious disease outcomes, to road safety measures. Data can be downloaded for one or many countries over multiple years.

Owner	World Health Organization
Website	https://www.who.int/data/gho
Coverage	National, Global

How to Access

Navigate the dashboard site to the indicator you want to explore. You can download data in a CSV format.

STATE OF GLOBAL AIR

This site focuses on air quality and health impacts of air pollution over time, by country.

Owner	State of Global Air
Website	https://www.stateofglobalair.org/data/#/air/plot
Coverage	National, Global

How to Access

Navigate the dashboard site to the indicator you want to explore. You can download data in a CSV format.

IV. SPATIAL VECTOR DATA

INTRODUCTION

Data from Sections II and III are mainly stored in raster grid style formats. Geospatial data stored in vector formats describe discrete objects on the Earth's surface. Information about roads, rivers, lakes, healthcare facilities, cellphone towers, and other points of interest can be stored in vector formats along with their location on the Earth's surface. This section describes vector data that our team has collected so far.

ADMINISTRATIVE BOUNDARIES

Administrative boundaries can be used for data visualization and analysis. The following table details what administrative boundary vector data we currently have and the last year the file was updated.

Country	Level	Year	Source	File Name
Malawi	Region	2003	FEWSNET	MW_Admin1_2003
	District	2003	FEWSNET	MW_Admin2_2003
	Country	2018	GADM	gadm36_MWI_0
	District	2018	GADM	gadm36_MWI_1
	Town/City/Traditional Authorities	2018	GADM	gadm36_MWI_2
	Unspecified	2018	GADM	gadm36_MWI_3
Zimbabwe	Province	2011	FEWSNET	ZW_Admin1_2011
	District	2011	FEWSNET	ZW_Admin2_2011
	Country	2018	GADM	gadm36_ZWE_0
	Province	2018	GADM	gadm36_ZWE_1
	District	2018	GADM	gadm36_ZWE_2
Gabon	Country	2018	GADM	gadm36_GAB_0
	Province	2018	GADM	gadm36_GAB_1
	Department	2018	GADM	gadm36_GAB_2
Zambia	Province	2012	FEWSNET	ZM_Admin1_2012
	District	2012	FEWSNET	ZM_Admin2_2012
	Country	2018	GADM	gadm36_ZMB_0
	Province	2018	GADM	gadm36_ZMB_1
	District	2018	GADM	gadm36_ZMB_2
Ghana	Country	2018	GADM	gadm36_GHA_0
	Region	2018	GADM	gadm36_GHA_1
	District	2018	GADM	gadm36_GHA_2
DRC	Province	2015	FEWSNET	CD_Admin1_2015
	Territory	2015	FEWSNET	CD_Admin2_2015
	Country	2018	GADM	gadm36_COD_0
	Province	2018	GADM	gadm36_COD_1
	Territory/Town	2018	GADM	gadm36_COD_2

Sources

FEWSNET (Famine Early Warning Systems Network) and GADM (Global Administrative Areas Database) maintain public use datasets of administrative boundaries for countries globally. Some countries have a version from both sources for a certain administrative level because administrative boundaries change over time. Some survey datasets may have sampling schemes based on past administrative boundaries.

How to Access

Download the data directly from the links in the Source column.

CELL PHONE TOWERS

OPENCELLID

OpenCellID is an open source database collecting information about where the radio signal of a GSM base station has been received. Using this data, the locations of cellphone towers are approximated. Point locations for the cellphone towers are included in this dataset.

Owner	OpenCellID
Website	https://opencellid.org/#zoom=16&lat=37.77889&lon=-122.41942
Coverage	Global

How to Access

Register an account at https://opencellid.org/register.php to receive an API access token. On the homepage, navigate to the "Data" tab and download data using the token.

ELECTRICAL GRID LINES

ENERGYDATA.INFO – ELECTRICAL GRID LINES

ENERGYDATA.INFO is an open data platform providing access to datasets and data analytics that are relevant to the energy sector. ENERGYDATA.INFO has been developed as a public good available to governments, development organizations, private sector, non-governmental organizations, academia, civil society and individuals to share data and analytics that can help achieving the United Nations' Sustainable Development Goal 7 of ensuring access to affordable, reliable, sustainable and modern energy for all. The platform is an innovation of the World Bank Group.

The Energy Transmission Network vector datasets are available for Malawi, Zambia, Zimbabwe, DRC, Gabon, and Ghana. Vectors are gathered from OpenStreetMap, the Africa Infrastructure Country Diagnostic dataset (World Bank), ECREEE transmission network for West Africa (online at ECOWREX), and other country-specific sources, then combined to include a country-wide geospatially located set of the main electricity grid lines.

Owner	ENERGYDATA.INFO
Website	https://energydata.info/
Coverage	National

How to Access

Search "Electricity Transmission Network" and the country name at https://energydata.info/dataset. Download directly from the site to the appropriate format.

HEALTH FACILITIES

MAINA ET AL. 2019 – PUBLIC HEALTH SECTOR FACILITIES

Maina and colleagues (2019) assembled a national master health facility lists from a variety of government and non-government sources from 50 countries and islands in sub-Saharan Africa and used multiple geocoding methods to provide a comprehensive spatial inventory of 98,745 public health facilities. Due to difficulties in obtaining complete data for publicly managed health care facilities, the authors focused on providers of public health sector services that cover services including expanded immunization programmes, health data surveillance and receive government funding to provide services to the general population.

This dataset is nationally comprehensive and locations only.

Owner	Maina et al.. (2019)
Website	https://www.nature.com/articles/s41597-019-0142-2
Coverage	National

Reference:

Maina, J., Ouma, P.O., Macharia, P.M. *et al.* A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Sci Data* **6,** 134 (2019).

How to Access

Link to WHO platform for downloading the data in *Nature* article is broken. Contact the authors of the article to ask for access to the data.

SERVICE PROVISION ASSESSMENTS (SPA) – DEMOGRAPHIC & HEALTH SURVEYS

The SPA surveys are administered by USAID to assess healthcare facilities' ability to provide services. They are administered to representatively sampled for national coverage and include private and public health facilities; pharmacies and individual doctors' offices are excluded. Typically, 400-700 facilities are included in a survey. Health care providers are interviewed to answer questions about facility infrastructure, supplies, medicines, vaccines, staff training and qualifications, and diagnostic capabilities.

GPS coordinates for health facilities are included.

Owner	USAID
Website	https://dhsprogram.com/Methodology/Survey-Types/SPA.cfm
Coverage	National

How to Access

Register an account on the DHS site (https://dhsprogram.com/data/new-user-registration.cfm). With a registered account, you may download any of the publicly available datasets.

HYDROLOGICAL FEATURES

INLAND WATER LINE AND WATER BODIES

These are vectors of lakes, streams, and shorelines for a given country. These vector data are useful for data visualization and calculating distance to water for spatial analysis. The data were downloaded from OpenStreetMap, a crowd-sourced database including vector datasets for land features, infrastructure, and points of interest.

Owner	OpenStreetMap
Website	https://www.openstreetmap.org
Coverage	National

How to Access

OpenStreetMap is being updated constantly. Download all vector data for a country from the Geofabrik Download Server at https://download.geofabrik.de/ for the most up to date data. The downloaded ZIP folder will contain all vector data for a selected country or region.

ROADS

OPEN STREET MAP ROADS

These are vector of roads for a given country. The data were downloaded from OpenStreetMap, a crowd-sourced database including vector datasets for land features, infrastructure, and points of interest. OpenStreetMap data are used when there is no comprehensive dataset for roads available. Because OpenStreetMap is a crowd-sourced dataset, it may be incomplete for some areas.

Owner	OpenStreetMap
Website	https://www.openstreetmap.org
Coverage	National

How to Access

OpenStreetMap is being updated constantly. Download all vector data for a country from the Geofabrik Download Server at https://download.geofabrik.de/ for the most up to date data. The downloaded ZIP folder will contain all vector data for a selected country or region.