THE ROBERT S. STRAUSS CENTER™
FOR INTERNATIONAL SECURITY AND LAW

CCAPS

CLIMATE CHANGE
AND AFRICAN
POLITICAL STABILITY

Sub-national African Education and Infrastructure Access Data

August 19, 2013

Todd G. Smith, Joshua W. Busby and Anustubh Agnihotri

LBJ School of Public Affairs, The University of Texas at Austin

August 19, 2013

Sub-national African Education and Infrastructure Access Data

Todd G. Smith, Joshua W. Busby and Anustubh Agnihotri
LBJ School of Public Affairs, The University of Texas at Austin
Corresponding Author: Todd G. Smith (toddgsmith@utexas.edu)

This dataset provides data on literacy rates, primary and secondary school attendance rates, access to improved water and sanitation, household access to electricity, and household ownership of radio and television. Unlike other datasets, notably the World Bank's World Development Indicators (WDI), this dataset provides data at the subnational level, specifically the first administrative district level. Furthermore, the data is comparable both within and across countries. This subnational level of data allows for assessment of education and household characteristics at a more relevant level for allocation of resources and targeting development interventions.

This data was calculated using raw survey data from three sources: the Demographic and Health Surveys (DHS) supported by the U.S. Agency for International Development (USAID)[1]; Multiple Indicator Cluster Survey (MICS) supported by UNICEF[2]; the General Household Surveys conducted by Statistics South Africa.[3] The datasets used are freely available for download from the websites of these agencies.

It is possible for one to extract the sub-national indicators directly from the original survey data and to request some of the data from DHS already aggregated at the sub-national level. That being said, the value-added of this dataset is fourfold. *First*, we combine all the publicly available sub-national data for these indicators from different sources into a single dataset. *Second*, we have connected these indicators with a unique identifier to shape files with the most current (2013) subnational units for the entire continent of Africa. *Third*, while the DHS regularly disseminate sub-national data on a variety of indicators, they have not done so for the specific education indicators in this dataset. *Finally*, for several of the indicators, including access to improved water sources and school attendance, we aggregate and report the indicators to allow for comparisons across countries.

Future iterations of this dataset will hopefully be improved in two significant ways. First, we hope to make this a panel dataset by including earlier versions of DHS data. Second, we hope to include more countries for which survey data will become available or is already available but for earlier years.

---

[1] http://www.measuredhs.org
[2] http://www.childinfo.org/mics
[3] http://www.statssa.gov.za

The final dataset includes data for 38 countries, covering 471 of Africa's 699 first level administrative districts.

NOTE: An earlier version of the dataset required users to perform the join in ArcGIS between published shape files and data files. This version includes shape files with the data already joined. For those administrative regions with missing data, this is indicated with a value of -99999. The user must exclude these values before visualizing or analyzing the data.

SUB-NATIONAL ADMINISTRATIVE REGIONS

Creating shapefiles

The final dataset includes data for 471 first-level administrative districts in 38 countries. Sao Tome and Principe with two principal islands is the country with the fewest, while Nigeria has the most at 37 states. For mapping purposes we sought to create a set of the most up-to-date administrative divisions using data from three freely available sources, none of which is completely up-to-date on its own. These sources are the Global Administrative Areas (GADM), Global Administrative Unit Layers (GAUL), and Map Library. In some cases, these sources were supplemented with boundary files from local governments in Africa. When evaluating the available sources we applied the following guidelines for deciding which of the sources to use for each country.

1. *Most up-to-date administrative borders.* For example, GADM and GAUL provincial boundaries for South Africa are out of date. We chose to use Map Library because it is more current. Still, the borders for Gauteng province did not reflect a recent change. We, therefore, updated these shapefiles with one obtained from Ekurhuleni Municipality.
2. *Shapes that match DHS regions the best.* For example, in Malawi, we use Map Library's three large regions as these reflect the same boundaries of the DHS regions.
3. *If more than one was up to date, we used the one that had the most detailed and geographically accurate borders or whose outside borders best matched neighboring countries.* For example, we use Map Library shapefiles for countries bordering South Africa – Namibia, Lesotho, Swaziland, Zimbabwe, Mozambique – because they match the South African borders.

Finally, we made slight adjustments to the national borders to ensure that neighboring countries matched exactly.

Survey sampling

DHS employs a stratified sampling strategy. Each household appearing in the DHS surveys is assigned to a *cluster*, which is used as the primary sampling unit. Clusters generally correspond to villages or towns in rural areas or neighborhoods in urban areas. These clusters are then grouped into a secondary sampling unit or *strata*. In many cases, the sampling strata employed

by DHS correspond to national administrative boundaries, which, in turn, correspond to our chosen shapefiles regions. In such cases the *strata* (region) variable reported in the survey data (generally v024) is used to identify sub-national regions and calculate the variables accordingly.

In some cases, however, DHS strata do not correspond to national geographic boundaries or GADM regions. In such cases, it is necessary to aggregate the data differently than it is presented in the DHS reports. In cases where the DHS region boundaries follow national administrative boundaries but include several sub-national divisions, the calculated indicator for the larger DHS region is simply applied to each administrative division included in the DHS region. This is the case for DRC, Côte d'Ivoire, Djibouti, Guinea-Bissau, and Republic of Congo. This approach has the benefit of allowing us to use the readily available shapefiles and potentially apply additional data to smaller units, but it has the potential to introduce bias due to the artificial inflation of the number of administrative divisions with distinct data. Anyone wishing to use this data for statistical analysis should pay careful attention to this issue and take steps to minimize any potential bias.

UNICEF's MICS employ a similar sampling strategy. Although not as stratified, the survey data identifies the region of each respondent household (generally variable hh7). Furthermore, MICS is designed to produce data comparable to that of DHS. The MICS website contains the following statement:

> UNICEF works closely with other household survey programmes, in particular the Demographic and Health Surveys (DHS) programme, to harmonize survey questions and modules and to ensure a coordinated approach to survey implementation, with the objective to provide comparability across surveys and to avoid duplication of efforts.[4]

Table 1 describes the set of countries covered, the source of the shapes, the number of sub-national administrative regions, and the survey source and year for the data.

Table 1: Country characteristics

Country Name	ISO Code	Shape Source	Regions	Survey Source	Survey Year º
Benin	BEN	GADM	12	DHS	2006
Burkina Faso	BFA	Map Library	14	DHS	2010
Burundi	BDI	Map Library	17	DHS	2011
Cameroon	CMR	GADM	12	DHS	2011
Central African Republic	CAF	GADM	17	MICS	2006

[4] "UNICEF - Statistics and Monitoring - Multiple Indicator Cluster Survey (MICS)," *UNICEF*, May 25, 2012, http://www.unicef.org/statistics/index_24302.html, accessed July 5, 2013.

Congo, Dem. Rep. of	COD	Map Library	26	DHS	2007
Congo, Republic of	COG	GADM	12	DHS	2005
Côte d'Ivoire	CIV	GADM	21	MICS	2006
Djibouti	DJI	GADM	5	MICS	2006
Egypt	EGY	GADM	26	DHS	2008
Ethiopia	ETH	GADM	11	DHS	2011
The Gambia	GMB	GADM	6	MICS	2006
Ghana	GHA	GAUL	10	DHS	2008
Guinea	GIN	GAUL	8	DHS	2005
Guinea Bissau	GNB	GADM	9	MICS	2006
Kenya	KEN	GADM	8	DHS	2009
Lesotho	LSO	Map Library	10	DHS	2010
Liberia	LBR	GADM	15	DHS	2007
Madagascar	MDG	GADM	22	DHS	2009
Malawi	MWI	Map Library	3	DHS	2010
Mali	MLI	GADM	9	DHS	2006
Mauritania	MRT	GADM	13	MICS	2007
Mozambique	MOZ	GAUL	10	DHS	2003
Namibia	NAM	Map Library	13	DHS	2007
Niger	NER	GADM	8	DHS	2006
Nigeria	NGA	GADM	37	DHS	2008
Rwanda	RWA	Map Library	5	DHS	2011
Sao Tome and Principe	STP	GADM	2	DHS	2009
Senegal	SEN	GADM	14	DHS	2011
Sierra Leone	SLE	GADM	4	DHS	2008

Somalia	SOM	GADM	18	MICS	2010
South Africa	ZAF	Map Library*	9	StatsSA	2011
Swaziland	SWZ	Map Library	4	DHS	2007
Tanzania	TZA	Map Library	26	DHS	2010
Togo	TGO	GADM	6	MICS	2006
Uganda	UGA	DHS	10	DHS	2006
Zambia	ZMB	GADM	9	DHS	2007
Zimbabwe	ZWE	Map Library	10	DHS	2011

º – If the survey fieldwork spanned the end of one year and the beginning of another, the latest year is reported.
* – Map Library shapes are modified slightly using shapefiles obtained from Ekurhuleni Municipality to reflect recent changes in municipal borders.

INDICATORS

The primary variables included in this dataset are listed in Table 1 below. Additionally, variables are included for the standard error, denoted with the suffix "_se" after the variable name, and sample size, denoted with suffix "_sz", to allow users to account for uncertainty in the data. Appendix A provides the summary statistics for the indicators. Appendices B-D provide exemplar do files for Ethiopia for 2011 for extracting the indicators. Contact the authors if you wish do files for any other countries.

Table 2: Variable names and descriptions

Variable name	Variable label
name_0	Country name
iso	ISO code
iso_num	ISO numeric code
name_1	Subnational region name
ccaps_id	CCAPS subnational region ID
year	Year

source	Data Source
website	Website
pnar	Net primary attendance rate (%)
pgar	Gross primary attendance rate (%)
snar	Net secondary attendance rate (%)
sgar	Gross secondary attendance rate (%)
lit15up	Literacy rate (15 & over)
lit1524	Literacy rate (15-24)
lit1519	Literacy rate (15-19)
lit2549	Literacy rate (25-49)
iwatpop	Access to improved water (% of population)
isanpop	Access to improved sanitation (% of population)
elecpop	Electricity in household (% of population)
radpop	Radio in household (% of population)
tvpop	Television in household (% of population)
rdtvpop	Radio and/or Television in household (% of population)
iwathh	Access to improved water (% of households)
isanhh	Access to improved sanitation (% of households)
elechh	Electricity in household (% of households)
radhh	Radio in household (% of households)
tvhh	Television in household (% of households)
rdtvhh	Radio and/or television in household (% of households)

School Attendance Rates

Net and gross enrollment rates are provided for primary and secondary school in each administrative region. The gross attendance rate is the total number of children attending primary or secondary school divided by the number of children within the official age range for

primary and secondary school. This indicator can potentially be greater than 100 because children outside of the official age ranges may be attending primary or secondary school. The net attendance rate is defined as the number of children between official primary or secondary school ages attending primary or secondary school divided by the total of number of primary or secondary school age children. This indicator is, by definition, bounded to 100 percent at the upper end.

Primary and secondary school attendance is determined by whether an individual household member is reported as attending school (v121) and in which education level (v122) and grade (v123) during the survey year. The age of the member is compared to official primary and secondary school grades in the country. Official primary and secondary grades and age ranges for each specific country are determined using information reported by the national statistical offices in the DHS reports or, in cases in which the information is unavailable or ambiguous from the DHS report, by information reported in UNESCO country profiles.

Literacy Rates

We report literacy rates for the following age groups: 15 and over; 15 to 24; 15 to 19; and 25 to 49. An individual is considered *literate* if they can read parts of or a whole sentence (v155) or if they have completed at least some secondary school (v149).

Water and Sanitation

DHS provides data at the household level on the source of drinking water (v201) and the type of toilet facility used (v205). Using criteria from the WHO / UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation,[5] these variables are recoded into *improved* water and *improved* sanitation. For example, piped water, protected wells and rainwater are all considered improved, while unprotected wells, surface water and water purchased from cart vendors is considered not improved. In accordance with guidelines of the JMP, if the household reports using bottled water for drinking, the source of water for other purposes (v202) is used to determine whether the drinking water is considered improved. Improved sanitation facilities include flush toilets, pit latrines and composting toilets. Specifics about this recode can be found in the Stata programs for each country.

Electricity, Radio and Television

Indicators are included for access to electricity, radio and/or television, both as a percentage of total households and as a percentage of total population. In both cases, access is determined by the presence of electricity (v206), radio (v207) and/or television (v208) in the household. Nonetheless, different indicators may result in different values depending on the size of the households in the region.

[5] WHO and UNICEF, "WHO / UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation," http://www.wssinfo.org/definitions-methods/watsan-categories.

COUNTRY SPECIFIC INFORMATION AND AGGREGATION

A discussion of the sub-national aggregation of each country as well as specific enrollment rates is provided below.

Benin
In Benin the DHS uses 23 strata but identifies the 12 official provinces as their reporting regions. We use the 12 official provinces from DHS and apply these to province shapefiles from GADM. No further aggregation is necessary. Primary school is grades 1 through 6 and secondary school includes 7 grades. Official primary school age is 6–11 and secondary school age is 12–18.

Burkina Faso
DHS divides Burkina Faso into 14 regions including a separate region for the capital, Ouagadougou. We assign these same aggregations to first administrative level shapefiles from Map Library for the 13 provinces plus a second administrative level shapefile for the capital city. In Burkina Faso, the first seven grades constitute the primary level and the next seven are the secondary level. Official primary school age is 7–12 and secondary school age is 13–19.

Burundi
DHS stratifies its sample using 17 regions. We assign data aggregations for these regions to corresponding Map Library shapefiles. In Burundi primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 12–17.

Cameroon
DHS reports for 12 regions in Cameroon: the ten provinces plus the cities of Yaounde and Douala. We apply the DHS aggregations to the 10 first administrative divisions from GADM as well as the second administrative division shapefiles for Yaounde and Douala. In Cameroon primary school is six years and secondary school is seven years. Official primary school age is 6–11 and secondary school age is 12–18.

Central African Republic (CAF)
For CAF we use MICS data, which stratifies the sample by 17 first administrative divisions. We apply aggregated data to the GADM shapefiles for the same 17 first administrative units. In CAR primary school is six years and secondary school is seven years. Official primary school age is 6–11 and secondary school age is 12–18.

Congo, Democratic Republic of
DHS stratifies DRC into 10 regions. We apply aggregated indicators for these 10 regions to 26 first administrative divisions obtained from Map Library, each of which is located entirely within one of the 10 DHS regions. In DRC primary school is six years and secondary school is seven years. Official primary school age is 6–11 and secondary school age is 12–18.

Congo, Republic of
In the Republic of Congo, DHS uses four regions: *nord*, comprised of the five northern departments; *sud*, including five southern departments; the city of *Brazzaville*; and the city of *Pointe Noire*. We apply the aggregated data for the two cities, Brazzaville and Pointe Noire, to second-level administrative division shapefiles from GADM. We apply the aggregated data for the nord and sud regions to the corresponding first level administrative shapefiles for the provinces included in each region. According to UNESCO, primary school in Congo is six years, official ages 6–11, and secondary school is seven years, official ages 12–18.

Côte d'Ivoire
MICS aggregates using 11 regions, 10 regions plus the city of Abidjan. We apply indicator data from these regions to the 21 first administrative divisions shapefiles obtained from GADM**.** In Cote d'Ivoire primary school is six years and secondary school is seven years. Official primary school age is 6–11 and secondary school age is 12–18.

Djibouti
MICS stratifies the survey into only two divisions: Djibouti City and the other areas. GADM includes shapefiles for five first administrative divisions. We apply the indicator data from the areas outside of Djibouti City to the corresponding four GADM divisions. In Djibouti primary school is six years and secondary school is seven years. Official primary school age is 6–11 and secondary school age is 12–18.

Egypt
DHS identifies the governorate of each respondent household in the survey. We use this to aggregate the indicator data and attach them to shapefiles from GADM for the 26 governorates. In Egypt primary school is six years and secondary school is six years. Official primary school age is 6–11 and secondary school age is 12–17.

Ethiopia
DHS stratifies its sample using the 11 official provinces. We assign data aggregations for these regions to corresponding GADM shapefiles. According to the 2011 DHS report, primary school is eight years and official primary school age is 7–14 while secondary school is four years and official secondary school age is 15–18.

The Gambia
DHS stratifies its sample according to the six official provinces. We assign data aggregations from these provinces to GADM shapefiles. In the Gambia primary school is six years and secondary school is six years. Official primary school age is 7–12 and secondary school age is 13–18.

Ghana
DHS stratifies its sample using the 10 official provinces. We apply these to shapefiles from GAUL, one of which is the province of Greater Accra. We, therefore, use the same aggregation

as DHS. Primary and secondary school in Ghana are both six years. Official primary school age is 6–11 while secondary school age is 12–17.

Guinea
DHS stratifies its sample using eight official regions. We assign data aggregations for these regions to corresponding GAUL shapefiles. Primary school in Guinea is six years, official ages 7–12, and secondary school is seven years, official ages 13–19.

Guinea-Bissau
MICS stratifies its sample using four regions: Est, Nord, Sud, and Bissau. We aggregate data according to these regions and then assign the calculated indicators to GADM shapefiles for nine provinces according to the MICS region of the province. In Guinea-Bissau primary school is six years and secondary school is five years. Official primary school age is 7–12 and secondary school age is 13–17.

Kenya
DHS stratifies its sample using the eight provinces of Kenya, one of which is the province of Nairobi. We apply data aggregations from these provinces to GADM shapefiles. Primary school in Kenya includes grades 1 through 8 and official primary school age is 6–13. Secondary school is 4 years the official age for which is 14–17.

Lesotho
There are 10 provinces in Lesotho and DHS stratifies its sample using these provinces. We apply data for these provinces to shapefiles obtained from Map Library. Primary school in Lesotho is seven years and secondary school is five years. Official primary school age is 6–12 while official secondary school age is 13–17.

Liberia
DHS stratifies its sample into 16 regions: the 15 official provinces plus Monrovia. We aggregate Monrovia households into South Central province before calculating the indicators and apply the aggregated data to 15 GADM shapefiles. In Liberia primary school and secondary school are both six years. Official primary school age is 6–11 and secondary school age is 12–17.

Madagascar
DHS stratifies its sample according to the 22 regions. These regions correspond to 22 second administrative divisions in GADM. We assign data aggregations to these 22 GADM shapefiles. In Madagascar primary school is five years and secondary school is seven years. Official primary school age is 6–10 and secondary school age is 11–17.

Malawi
DHS stratifies its sample into three regions: Central, Northern, and Southern. We assign aggregated indicators for these three regions to corresponding Map Library shapefiles. In Malawi primary school is eight years and secondary school is four years. Official primary school age is 6–13 and secondary school age is 14–17.

Mali
DHS stratifies its sample using the nine regions. We assign data aggregations for these regions to corresponding GADM shapefiles. In Mali primary school is six years and secondary school is six years. Official primary school age is 6–11 and secondary school age is 12–18.

Mauritania
MICS includes 13 regions for surveyed households. We aggregate data to these 13 regions and assign the indicators to corresponding shapefiles from GADM. In Mauritania primary school is 6 years and secondary school is 7 years. Official primary school age is 6–11 and secondary school age is 12–18.

Mozambique
DHS stratifies its sample using the ten official provinces plus Maputo City. We assign Maputo City households to Maputo Province before calculating the indicators and assign data aggregations to ten corresponding GAUL shapefiles. In Mozambique primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 13–17.

Namibia
DHS stratifies its sample using the 13 official provinces. We assign data aggregations for these regions to corresponding Map Library shapefiles. In Namibia primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 13–17.

Niger
DHS stratifies its sample using eight regions. We assign data aggregations for these regions to corresponding GADM shapefiles. In Niger primary school is six years and secondary school is seven years. Official primary school age is 7–12 and secondary school age is 13–19.

Nigeria
We use the DHS GIS data to identify the state of each household and aggregate the indicators according to these 37 states. We then assign the aggregated data to 37 corresponding GADM shapes. In Nigeria primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 13–17.

Rwanda
DHS stratifies its sample using five regions. We assign data aggregations for these regions to corresponding Map Library shapefiles. Primary school is grades 1 through 6 and secondary school includes 7 grades. Primary school age is 7–12 and secondary school age is 13–18.

Sao Tome and Principe
DHS stratifies its sample using two regions corresponding to the two main islands. We assign data aggregations for these regions to corresponding GADM shapefiles. In Sao Tome and

Principe primary school is six years and secondary school is six years. Official primary school age is 6–11 and secondary school age is 12–17.

Senegal
In 2011, DHS stratified its sample using the 14 official provinces. This includes three previously second-level administrative divisions that have been elevated to the status of provinces since the previous survey: Kaffrine, Kedougou, and Sedhiou. We assign data aggregations for these regions to the 11 corresponding GADM first level administrative division shapefiles plus three second-level administrative shapefiles for the three new provinces. According to the 2010-2011 DHS report, Senegal primary school is six years and secondary school is five years. Official primary school age is 6–11 and secondary school age is 12–16. (This is a change from the report of the previous DHS survey round.)

Sierra Leone
DHS stratifies its sample using four regions: Eastern, Western, Northern, and Southern. We assign data aggregations for these regions to corresponding GADM shapefiles. In Sierra Leone primary school is six years and secondary school is six years. Official primary school age is 6–11 and secondary school age is 12–17.

Somalia
MICS includes 18 regions for surveyed households. We aggregate data to these 18 regions and assign the indicators to corresponding shapefiles from GADM. In Somalia primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 13–17.

South Africa
Stats SA collects survey responses according to the nine official provinces. We assign aggregated data to nine corresponding Map Library shapefiles modified slightly using shapefiles obtained from Ekurhuleni Municipality to reflect recent changes in municipal borders. In South Africa primary school is seven years and secondary school is five years. Official primary school age is 7–13 and secondary school age is 14–18.

Swaziland
DHS stratifies its sample using the four official provinces. We assign data aggregations for these regions to corresponding Map Library shapefiles. In Swaziland primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 13–17.

Tanzania
DHS stratifies its sample using 26 regions. We assign data aggregations for these regions to corresponding Map Library shapefiles. In Tanzania primary school is seven years and secondary school is six years. Official primary school age is 6–11 and secondary school age is 12–18.

Togo
MICS includes six regions for surveyed households. We use a GADM second-level administrative shapefile for Lome and add this to the five first-level administrative shapes. We then aggregate data to these six regions and assign the indicators to the corresponding shapefiles from GADM. In Togo primary school is six years and secondary school is seven years. Official primary school age is 7–13 and secondary school age is 14–19.

Uganda
DHS stratifies its sample using ten regions. We assign data aggregations for these regions to corresponding shapefiles obtained directly from DHS. In Uganda primary school is seven years and secondary school is seven years. Official primary school age is 6–12 and secondary school age is 13–18.

Zambia
DHS stratifies its sample using the nine official provinces. We assign data aggregations for these regions to corresponding GADM shapefiles. In Zambia primary school is seven years and secondary school is five years. Official primary school age is 6–12 and secondary school age is 13–17.

Zimbabwe
DHS stratifies its sample using the 11 official provinces. We assign data aggregations for these regions to corresponding Map Library shapefiles. In Zimbabwe primary school is six years and secondary school is six years. Official primary school age is 7–12 and secondary school age is 13–18.

THE ROBERT S. STRAUSS CENTER™
FOR INTERNATIONAL SECURITY AND LAW

CLIMATE CHANGE
AND AFRICAN
POLITICAL STABILITY

THE UNIVERSITY OF TEXAS AT AUSTIN

2315 RED RIVER STREET

AUSTIN, TEXAS 78712

PHONE: 512-471-6267

FAX: 512-471-6961

INFO@STRAUSSCENTER.ORG

CCAPS.STRAUSSCENTER.ORG

This material is based upon work supported by, or in part by, the U.S. Army Research Office contract/grant number W911NF-09-1-007 under the Minerva Initiative of the U.S. Department of Defense.