# Codebook and Data Collection Protocol

July 2018

Michaël Aklin, University of Pittsburgh and ISEP
S.P. Harish, William & Mary and ISEP
Johannes Urpelainen, Johns Hopkins University and ISEP

## General Comments

This document contains the data collection protocol and the codebook for the new database on household electrification collected by Michaël Aklin, S.P. Harish, and Johannes Urpelainen. For details, see:

> Aklin, Michaël, S.P. Harish, and Johannes Urpelainen. "A Global Analysis of Progress in Household Electrification." *Energy Policy*.

This project aimed at collecting reliable data from a range of sources on the following three variables:

a) Total Electrification Rate: percentage of total households in the country who have access to grid electricity

b) Rural Electrification Rate: percentage of rural households in the country who have access to grid electricity

c) Urban Electrification Rate: percentage of urban households in the country who have access to grid electricity

The data are available on Harvard Dataverse. The dataset is a file called *Electrification_Database.dta*.

The data were collected by Michaël Aklin, S.P. Harish, and Johannes Urpelainen. They gratefully acknowledge the financial support of the International Growth Centre (Grant E-89301-NOC-1).

# Protocol

The data were collected based on the following rules.

1. Research assistants (RA) were assigned to a specific country. Their primary task was to find data for the three variables listed above going as far back in time as feasible.
2. Sources: we provided RAs with the following guidelines to search for data, with the top priorities in **bold font** and sources not to use in **red bold font**:
   2.1. **National census reports: typically google "[country name] census [year]" to access reports on the census.**

2.2. **Nationally representative socio-economic surveys: typically google "[country name] socio economic survey [year]" or "[country name] household budget survey [year]"**

2.3. Department of Statistics website from the country

2.4. Ministry of Energy or Rural Electrification Authority of the country

2.5. Countries' utilities (especially when they are public monopolies).

2.6. Secondary literature (country case studies, etc. via Google Scholar) from history and from energy studies. Includes journals such as *Energy*, *Energy for sustainable Development*, *Energy Policy*, *Renewable Energy*, *Utilities Policy*, *World Bank Economic Review*, etc. Exception: sources that are not accepted (see next point).

**2.7. The following sources were not be used: WB, EIA, IEA, WHO, or IRENA, or any secondary material that uses these as data sources.**

3. Each observation was added to a spreadsheet. The source was documented and a copy of the original source was kept on an online server (generally in a pdf format).
4. Once a country was completed, each observation was reviewed by one of the PIs. This included vetting the source of the data.
5. Once all data for all countries were collected, the PIs reviewed each observation to detect unreliable observations. For instance, this included flagging sudden increases or decreases, or abnormally high/low values. In such cases, the PIs returned to the original source and verified the quality of the data. If needed, they searched for additional sources corroborating the validity of the data. If the observation was deemed unreliable, it was discarded.

## Codebook

| Variable Name | Description |
|---|---|
| **countryname** | Name of the country, following World Bank nomenclature. |
| **countrycode** | ISO3 country code. |
| **year** | Year of the observation. |
| **elecrate_total** | Total household electrification rate. |
| **elecrate_urban** | Urban household electrification rate. |
| **elecrate_rural** | Rural household electrification rate. |

## Notes

- The data rely almost exclusively on censuses and representative surveys. We only imputed cases in

situations in which we observed two out of the three variables and could back out the last one. For instance, when we had total electrification and rural electrification rates, and we also had data on the share of the population living in urban and rural areas, we could compute electrification rates in urban areas. This differs from traditional imputation because it uses population data to reconstruct the data.

- All observations are documented in *DataReferences.pdf*. For access to the raw data sources, please contact one of the authors.
- The list of countries we studied was based on the following criteria: it had to have more than 300,000 inhabitants and have had a reasonably stable and open political system. Countries that were war-torn (e.g. Afghanistan) or closed (e.g., North Korea) were therefore generally excluded, except if reliable data could be found.
- Electrification is defined with respect to households (% of households with access to electricity). In some cases, only population shares were available (% of total population with access to electricity). We verified that using the latter did not yield large differences with a test in India (for which we have both data at the household and the population level). The difference was small. See manuscript, footnote #2.
- We focused on *grid* electricity. Thus, we typically excluded households that only benefitted from solar microgrids or similar distributed power systems.
- When multiple sources listed different electrification rates, we used the more credible and nationally representative source.