

All data extracted from its original source is structured and standardised according to the NDAP standardisation process before it is hosted on the platform. Standardisation allows all datasets on NDAP to be reported using common administrative and temporal dimensions, ensuring that all datasets are comparable and interoperable across the platform. To read more about this standardisation process, see the documentation hosted on NDAP: [Click Here](#)

When exporting a dataset from NDAP, not only is the final, standardised data file downloaded, but all of the files required to apply the NDAP standardisation process are included in the download as well. This zip folder contains the raw data, the standardised data, and the files required to replicate the standardisation. The contents of this folder includes:

(Note that 7049 indicates the unique 4 digit source code for this dataset, each dataset will have it's own unique source code).

- 7049\_source\_data.csv: The raw data from the source with original administrative dimensions. This dataset may have already been restructured by scraping PDFs, combining files, or pivoting tables to fit the proper tabular format used by NDAP, but the actual data values remain unchanged.
- NDAP\_REPORT\_7049.csv: The final standardised data using LGD geographic dimensions as seen on NPAP.
- 7049\_metadata.csv: Variable-level metadata, including the following fields:
  - ❖ VariableName: The full variable name as it appears in the data
  - ❖ VariableCode: A unique variable code that is used as a short name for the variable during internal processing and can be used for simplicity if desired
  - ❖ Type\_Of\_Variable: The classification of the column, whether it is a dimension or a variable (i.e. indicator)
  - ❖ Unit\_Of\_Measure:
  - ❖ Aggregation\_Type: The default aggregation function to be used when aggregating each variable
  - ❖ Weighing\_Variable\_Name: The weight assigned to each variable that is used by default when aggregating
  - ❖ Weighing\_Variable\_ID: The weighting variable id corresponding to the weighing variable name
  - ❖ Long\_Description: A more descriptive definition of the variable
  - ❖ Scaling\_factor: Scaling factor from source
- 7049\_KEYS.csv: The key which maps source administrative units to the standardised Local Government Directory (LGD) dimensions. This file also contains pre-calculated weights for every constituent unit mapped from the source dimensions into the LGD. You can interpret each row as describing what fraction of the source unit is mapped to a corresponding LGD unit. This file includes the following fields:
  - ❖ src[Unit]Name: The administrative unit name as it appears in the source data. Depending on the dataset, that may include State, District, Subdistrict, Block, Village/Town, etc.
  - ❖ [Unit]Name: The standardised administrative unit name as it appears in the LGD. Depending on the dataset, that may include State, District, Subdistrict, Block, Village/Town, etc.
  - ❖ [Unit]Name: The standardised administrative unit code corresponding to the unit name in the LDG.
  - ❖ Year: The year in which the data was collected or reported. Depending on the dataset, any other temporal variables may also be present (Quarter, Month, Calendar Day, etc.)
  - ❖ Number\_Of\_Children: The number of LGD units associated with the mapping described by an individual row. Units from the source that have undergone a split will contain multiple children.
  - ❖ Number\_Of\_Parents: The number of source units associated with the mapping described by an individual row. Units from the source that have undergone a merge will contain multiple parents.
  - ❖ Weighing\_Variables: Households, Population, Male Population, Female Population, Land Area (Total, Rural, and Urban versions of each). For each weighing variable there are the following associated fields:
    - Count: the total count of households, population, or land area mapped from the source unit to the LGD unit for that particular row (NumberOfHouseholds, TotalPopulation, LandArea).
    - Mapping\_Error: the percentage error due to missing villages in the base data, meaning what fraction of the weighing variable is dropped because the microdata could not be mapped to the LGD.
    - Weighing\_Ratio: the weighing ratio for that constituent match of source unit to LGD unit for each particular row. This is the fraction applied to the source data to achieve the LGD-standardised final data.

**\*\*Disclaimer:** NDAP's standardised version of this data source is included in this download.

Standardisation employs a statistical weighting methodology to transform the data from its raw, original form into the standardised units of the Local Government Directory (LGD), altering the data from its original form. Read more about this process in [NDAP's standardisation documentation](#) to fully understand the methodology. Please see here for the complete Legal disclaimer of using NDAP data.