Institute for Health Metrics and Evaluation

# Data Release Information Sheet

## *Data Summary*

Dataset name: Low- and Middle-Income Country Drinking Water and Sanitation Facilities Access Geospatial Estimates 2000-2017

Date of release: August 19, 2020

Summary:

Annual estimates were produced for access to drinking water and sanitation Facilities at the 5x5 km-level for 90 low- and middle-income countries (LMICs) for 2000-2017. These estimates were produced using a geo-positioned dataset created from 634 household surveys. Survey sources used include the Demographic and Health Survey (DHS) and UNICEF Multiple Indicator Cluster Survey (MICS) series, and other country-specific surveys.

This dataset includes the following:

- GeoTIFF raster files for pixel-level estimates of drinking water and sanitation facility coverage percent (percent of people with the given type of access) and number (number of people with the given type of access)
- CSV files of aggregated estimates for each country at zero, first and second administrative divisions
- Code files used to generate the estimates

Get Data Files

Relevant publications and visualizations:

- Deshpande A, Hay, SI, Reiner RC Jr., Local Burden of Disease WaSH collaborators. Mapping geographic inequalities in access to drinking water and sanitation facilities in low- and middle-income countries, 2000–2017. *The Lancet Global Health*. 19 Aug 2020.
- Local Burden of Disease – WASH

## *Acknowledgements*

Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)

Funders:

- Bill and Melinda Gates Foundation (BMGF)

Suggested Citation:

Institute for Health Metrics and Evaluation (IHME). Low- and Middle-Income Country Drinking Water and Sanitation Facilities Access Geospatial Estimates 2000-2017. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020.

## *Data Files Information*

### CSV files of Drinking Water and Sanitation Facility Coverage

Stored in files named <MEASURE>_<LEVEL_OF_AGGREGATION>.CSV

(Example: IHME_LMIC_WASH_2000_2017_**W_PIPED_ADMIN_1**_ Y2020M06D02.CSV)

- **Measure**: w_piped (access to piped water), w_imp (access to any improved water sources), w_imp_other (access to non-piped improved water sources), w_unimp (reliance on unimproved water sources), w_surface (reliance on surface water), s_piped (access to sewer and septic sanitation facilities), s_imp (access to any improved sanitation facility), s_imp_other (access to a non-piped improved sanitation facility), s_unimp (reliance on unimproved sanitation facilities), s_od (reliance on open defecation)
- **Level of aggregation:** admin0, admin1, or admin2, corresponding to zero, first, and second administrative level areas as defined in the Database of Global Administrative Areas (GADM) 2019 shapefiles, with adjustments made in some countries. Each row in each table is unique by administrative unit and year

| Variable | Variable Label | Variable Definition |
|---|---|---|
| ADM0_CODE | GADM Admin 0 Code | GADM code identifying the administrative unit |
| ADM0_NAME | Admin 0 Name | Zero level administrative unit (Country) name as found in the GADM shapefile |
| ADM1_CODE | GADM Admin 1 Code | GADM code identifying the administrative unit (Only in the admin1 files) |
| ADM1_NAME | Admin 1 Name | First level administrative unit name as found in the GADM shapefile |
| ADM2_CODE | GADM Admin 2 Code | GADM code identifying the administrative unit (Only in the admin2 files) |
| ADM2_NAME | Admin 2 Name | Second level administrative unit name as found in the GADM shapefile (Only in the admin2 files) |

| Variable | Variable Label | Variable Definition |
|---|---|---|
| year | Year | Time period of estimate. Possible values: years in the range 2000-2017 |
| age_group_id | Age Group ID | Unique numeric identifier for the age group generated and stored in an IHME database of data dimensions. Possible values: 22 |
| age_group_name | Age Group Name | Age group estimated. Possible values: All ages |
| sex_id | Sex ID | Unique numeric identifier for the sex generated and stored in an IHME database of data dimensions. Possible values: 3 |
| sex | Sex | Sex estimated: Possible values: Both |
| measure | Measure | The measure (indicator) estimated. Possible values: w_piped, w_imp, w_imp_other, w_unimp, w_surface, s_piped, s_imp, s_imp_other, s_unimp, s_od |
| metric | Metric | Metric/unit of measure for the estimate. Values: Percent and Number |
| mean | Mean | Mean posterior population-weighted estimate for the administrative unit |
| upper | Upper Confidence Interval | 97.5% population-weighted posterior quantile estimate for the administrative unit |
| lower | Lower Confidence Interval | 2.5% population-weighted posterior quantile estimate for the administrative unit |

## Codebooks

Variable names, labels, and value encoding for admin 0 files can be found in the machine-actionable codebook file **IHME_LMIC_WASH_2000_2017_CODEBOOK_ADMIN_0_Y2020M06D02.CSV**

Variable names, labels, and value encoding for admin 1 files can be found in the machine-actionable codebook file **IHME_LMIC_WASH_2000_2017_CODEBOOK_ADMIN_1_Y2020M06D02.CSV**

Variable names, labels, and value encoding for admin 2 files can be found in the machine-actionable **IHME_LMIC_WASH_2000_2017_CODEBOOK_ADMIN_2_Y2020M06D02.CSV**

## GeoTIFF Raster Files for Pixel-level Estimates of Drinking Water and Sanitation Facility Coverage

Stored in files named <MEASURE>_<METRIC>_<STAT>_<YEAR>.TIF

(Example: IHME_LMIC_WASH_2000_2017_**W_PIPED_PERCENT_MEAN_2003**_ Y2020M06D02.TIF)

- **Measure**: w_piped (access to piped water), w_imp (access to any improved water sources), w_imp_other (access to non-piped improved water sources), w_unimp (reliance on unimproved water sources), w_surface (reliance on surface water), s_piped (access to sewer and septic sanitation facilities), s_imp (access to any improved sanitation facility), s_imp_other (access to a

non-piped improved sanitation facility), s_unimp (reliance on unimproved sanitation facilities), s_od (reliance on open defecation)

- **Metric**: percent, number
- **Stat:** mean, upper, or lower summary statistics from the predictive posterior distribution at each pixel. Lower and upper correspond to 2.5% and 97.5% quantiles
- **Year**: From 2000 to 2017, corresponding to the time period of the estimate

Note that rasters mask (i.e., have NA values) for lakes and areas with low population (10 people per 1km and classified as barren/sparsely vegetated).

### Data Input Sources

This file contains relevant metadata about the input sources as suggested in the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER), a statement that promotes best practices in reporting health estimates.

Sanitation Sources

**IHME_LMIC_WASH_2000_2017_SANI_DATA_INPUT_SOURCES_Y2020M06D02.XLSX**

Water Sources

**IHME_LMIC_WASH_2000_2017_WATER_DATA_INPUT_SOURCES_Y2020M06D02.XLSX**

## *Additional Information*

### Terms and Conditions

http://www.healthdata.org/about/terms-and-conditions

### Contact information

To request further information about this dataset, please contact IHME:

Institute for Health Metrics and Evaluation
2301 Fifth Ave., Suite 600
Seattle, WA 98121
USA
Telephone: +1-206-897-2800
Fax: +1-206-897-2899
Email: data@healthdata.org
www.healthdata.org

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.