

Reproducibility Package for the Social Sustainability Global Database (SSGD)

Paola Ballon
Omar Alburqueque
Global Department for Social Development
World Bank

June 2025

Contents

Overview	1
Data Availability	2
Dependencies	3
Contents of the Package	4
Instructions	4
Output Details	5
Licensing and Citation	5
Contact	5

Overview

Social sustainability is defined as the condition in which individuals are included in development and trust that they and future generations will benefit from it. Based on this definition, the Social Sustainability Global Database (SSGD) compiles 128 indicators from 18 recognized sources, structured around four dimensions of social sustainability: social cohesion, inclusion, resilience, and process legitimacy. The database covers 222 countries and territories, provides disaggregation by sex, age, urban/rural location, ethnicity, and religion, and includes information at both national and subnational (ADM1) levels. Data are organized into two periods: 2015–2018 and 2019–2022.

This README provides essential instructions for replicating the SSGD (to access the database [click here](#)). The replication process involves using four master files in the Stata language. This package includes only the script files; users should consult the [guidelines download raw data.xlsx](#) file for instructions on how to obtain the raw data¹.

Data Availability

Except for the microdata collected from the Global Monitoring Database (GMD), all data used in this study are publicly available but not included in the reproducibility package. For guidelines on how to access each dataset, see the document “[guidelines_download_raw_data.xlsx](#)”.

1. Equality of Opportunity for Sexual and Gender Minorities - EQOSOGI

Name: Equality of Opportunity for Sexual and Gender Minorities - EQOSOGI

Source: World Bank

Indicator list:

Access to inclusive education score
Access to labor market score
Access to public services and social protections score
Civil and political inclusion score
Criminalization and SOGI score
Protection from hate crimes score

Access policy: Published with the package

2. Sovereign ESG Data Portal

Name: Sovereign ESG Data Portal

Source: World Bank

Indicator list:

CO2 emissions
Cooling degree days
Energy intensity level of primary energy
Forest area (% of land area)
Heating degree days
Individuals using the Internet (% of population)
Level of water stress
Net migration
Patent applications, residents
Political Stability and Absence of Violence/Terrorism
Ratio of female to male labor force participation rate (%) (modeled ILO estimate)
Regulatory Quality
Research and development expenditure (% of GDP)
School enrollment, primary and secondary (gross), gender parity index (GPI)
Tree cover loss
Unmet need for contraception (% of married women ages 15-49)

Access policy: Published with the package

3. Human Development Index (HDI)

Name: Human Development Index (HDI)

Source: United Nations Development Programme

Indicator list:

Human Development Index

Access policy: Published with the package

4. Worldwide Development Indicators

Name: Women, Business and the Law

Source: World Bank

Indicator list:

% of urban population
Fertility rate
GDP growth (annual %)
GDP per capita, PPP
Gini index (World Bank estimate)
Human Capital Index
Inequality of opportunities (by using the Human Capital Index)
Inequality of opportunities index
Log of GDP per capita, PPP
Multidimensional poverty headcount ratio
Poverty headcount ratio at national poverty lines
Shared prosperity
Shared prosperity premium

Access policy: Published with the package

Dependencies

- **Software:** Stata 16 MP (or any higher version).
- **Optional:** In some cases, raw data can be downloaded automatically using Stata. For such cases, it is recommended to use the datalibweb package (restricted to World Bank staff) and the wbopendata package (publicly available and installable via `ssc install wbopendata`). Further details can be found in the guidelines download raw data.xlsx file.

Contents of the Package

The replication package includes the following key components:

- **Master files:** Located in the master files folder. They must be executed in the specified order to reproduce the SSGD. See the Instructions section.
 - 1.ssgd_user profile.do – Sets the working directory for the replication process.
 - 2.ssgd_data gen.do – Generates all indicators from the processed raw data.
 - 3.ssgd_data proc.do – Organizes the indicators by wave and applies source prioritization rules.
 - 4.ssgd_data merge.do – Merges all processed data and appends metadata and labeling.
- **Script files:** Located in the scripts folder. These can be grouped into data generation scripts and data processing scripts.
 - **Data generation** – Use raw data to generate indicators and, when feasible, estimate values by population groups or subnational units. Examples: afrobarometer8.do, afrobarometer8_subnational.do.
 - **Data processing** – Use the outputs from the data generation block to classify information by waves or apply source prioritization rules. Examples: acled_data_processing.do, civicus_data_processing.do.
- **Support files:** Located in the support data folder. These include ancillary files used by some of the script files. They typically contain metadata to be displayed in the SSGD.
- **Data folders:** These include the raw data and proc data folders. Only the raw data folder may be modified by the user, provided its structure remains unchanged and modifications are limited to copying raw data files into the corresponding subfolders. Further details can be found in the guidelines download_raw_data.xlsx file.
- **Final database:** Located in the final data folder. Once all master files have been executed, the SSGD will be automatically generated and saved in this folder.

Instructions

1. Open the 1_ssgd_user profile.do master file and set the working directory according to your local environment. It will all sub dofiles: 2 ssgd_data gen.do, 3 ssgd_data proc.do and 4 ssgd_data merge.do
2. Upon completion, the final database will be available in the final_data folder.

Output Details

The SSGD will be stored in the final data folder under the name ssgd_v2_long. It is provided in both Stata and Excel formats and follows a long-format structure.

Licensing and Citation

This work is distributed under the Creative Commons Attribution 4.0 International license (CC BY 4.0).

Citation: P. Ballon and O. Alburquerque (2024). *Social Sustainability Global Database 2.0: A leap forward in data disaggregation*. The World Bank Group.

Contact

For inquiries, please contact Paola Ballon – pballon@worldbank.org and Omar Alburquerque – oalburquechav@worldbank.org