**Standards for the Social Sustainability Global Database (SSGD) v2.0 Replication: A comprehensive guide**

**Paola Ballon and Omar Alburqueque[[1]](#footnote-1)**

**The World Bank**

All script files used to develop the SSGD v2.0 are publicly available in the GitHub repository (see <https://github.com/Paola-Ballon/Social-Sustainability-Global-Database-v2.0>). We designed and organized such files to ensure the replicability of this database by anyone who has access to the World Bank Group's intranet. This section briefly discusses the files used to develop the SSGD v2.0 and presents guidelines for their proper use.

To ensure ease of replicability of the SSGD v2.0, we have coded seven master files in the R and Stata languages. The user only needs to work with these master files, which we classify into the following five blocks: 1) User profile, 2) Data download, 3) Generation of indicators, 4) Data processing, and 5) Data merge. Figure 1 shows a sequential diagram that summarizes all the blocks and lists the master files included in each block.

**Figure 1**: Blocks involved in the replicability of the SSGD v2.0

*Software requirements*

Before discussing every block of the replication process in detail, we provide a list of requirements every user must fulfill to replicate the SSGD v2.0 without problems. First, the user must have both Stata 16 MP and the latest version of R installed on their device. These are minimum requirements; for instance, a user could use Stata 17, but we strongly encourage using the MP version. Also, the user can choose any development environment software to access the R scripts; RStudio is a popular choice among R users. Second, the user must install the datalibweb and wbopendata packages in Stata before running any do file in the SSGD v2.0 repository. The datalibweb package enables the user to access harmonized household surveys from the World Bank’s microdata library, such as GMD (see <http://spqsapps.worldbank.org/qs/ECA/_layouts/15/WopiFrame.aspx?sourcedoc=/qs/ECA/SiteAssets/datalibweb_Guidelines_1page.pdf&action=default> for further information on the package). On the other hand, the wbopendata package enables the user to access over 3000 indicators from the World Bank databases (see <https://datahelpdesk.worldbank.org/knowledgebase/articles/889464-wbopendata-stata-module-to-access-world-bank-data> for further information on the package).

1. **User profile**

In the user profile block, the user must declare, for one time, the path or working directory where she decided to clone the repository. This way, the rest of the files will adapt to any user-defined folder. The master files present in this block are "ssgd\_v2\_user\_profile.R" and "ssgd\_v2\_user\_profile.do". It is essential to declare the same path or working directory in both master files before executing them. By default, the working directory for both files is "C:/Users/PC/Desktop/" but the user may change it.

1. **Data download**

The data download block consists of a semiautomatic process of obtaining raw data to construct the indicators included in the SSGD v2.0. In other words, it entails downloading microdata and other data files. The master files included in this block are "ssgd\_v2\_data\_download\_p1.R" and "ssgd\_v2\_data\_download\_p2.do", the user must execute them in the order described right after running the master files of the previous block.

We say the data download block is semiautomatic because we execute part of it automatically using the master files. In contrast, the rest must be done manually by the user. Except for GMD, all data sources offer data for free. However, in some cases, access to data requires registration or filling out a data request form, so it is not always possible to obtain the information through web scrapping or related computational procedures[[2]](#footnote-2).

Table 1 lists the databases used in the SSGD v2.0 and identifies those requiring manual data download. For such cases, the procedure involves the following steps: 1) access a particular URL (depending on the source), 2) register or fill out a data request form, 3) download the data, and 4) assign it to a folder contained in the local repository. The user is not required to create or modify any folder within the local repository since all folders are created automatically along with the execution of the master file "ssgd\_v2\_data\_download\_p1.R". All downloaded files are stored in the “raw\_data” folder. Section A1 in the appendix provides guidelines for each database where the user must download data manually.

**Table 1**: Access and download of SSGD v2.0 data inputs

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Free Access** | **Registration required** | **Download Type** |
| GMD | No | No | Automatic |
| AF | Yes | No | Automatic |
| AB | Yes | No | Automatic |
| ASB | Yes | Yes | Manual |
| LB | Yes | No | Manual |
| WVS | Yes | Yes | Manual |
| EVS | Yes | Yes | Manual |
| ACLED | Yes | Yes | Manual |
| CIVICUS | Yes | No | Included |
| FINDEX | Yes | No | Automatic |
| WDI | Yes | No | Automatic |
| WGI | Yes | No | Automatic |
| WJP | Yes | No | Automatic |
| EIU | Yes | No | Included |
| ESG | Yes | No | Automatic |
| WBL | Yes | No | Automatic |
| External: EQOSOGI | Yes | No | Included |
| External: UNDP | Yes | No | Automatic |

1. **Generation of indicators**

Once the user completes the semiautomatic data download process, she must execute the master file "ssgd\_v2\_gen\_indicators.do" file, which generates all the indicators in the SSGD v2.0. It is necessary to run all the previous steps correctly so that the master file can create the indicators properly and without execution errors. The user will notice that the “proc\_data” folder (within the local repository) serves to store every file generated by this master file. The “proc\_data” name indicates that it is processed data (unlike the "raw\_data" folder).

1. **Data processing**

After creating all indicators, we classify the information by wave (either wave 1 which corresponds to the period 2015-2018, or wave 2 which corresponds to the period 2019-2022). The master file "ssgd\_v2\_data\_proc.do" will perform this task. Likewise, this master file aims to establish "dominance relationships" between the different databases used since there are cases where the same indicator is present or can be extracted from more than one data source.

There are three dominance relationships established in SSGD v2.0. In the first place, the FINDEX database dominates other databases for the “ownban” indicator (% of people with a bank account) since the FINDEX database is global (includes almost all countries) and presents the information with fixed periodicity. Secondly, the AF database dominates the AB one only for the indicators they have in common and for the following countries: Morocco, Sudan, and Tunisia. The reason is AF provides more information in general, and the years included are more recent. Finally, barometers (AF, AB, ASB, and LB) dominate the WVS database in their common indicators and countries since the countries in the WVS database appear once every several years.

1. **Data Merge**

For the final block in the replication process of the SSGD v2.0, the user must run the “ssgd\_v2\_merge.do” to merge all the indicators processed so far. Unlike the processed data in the previous block, the "final\_data" folder will contain the final product of this master file, i.e., the data file "ssgd\_v\_2\_0.dta". Figure 2 presents a chronological diagram summarizing the steps to replicate the SSGD v2.0.

**Figure 2**: A diagram for the replication process of the SSGD v2.0



**Appendix**

Section A1: Instructions to manually download selected databases

Asianbarometer: The user must fill out a data request form by clicking on the following link: <https://www.asianbarometer.org/data?page=d10>. After that, the user must select all the databases required. For the case of ASB wave four, the user must choose the "All Country" option. For ASB wave 5, the user must select the data files for the following countries: Philippines, Mongolia, Australia, India, Indonesia, Japan, South Korea, Malaysia, Myanmar, Taiwan, Thailand, and Vietnam. After that, a download portal will open with a list of the selected data files, and the user must download and unzip all the files listed. Finally, the user must add the uncompressed data to the folders "asianbarometer4" and "asianbarometer5" appropriately. These folders are inside a folder called "raw\_data" which, in turn, is located inside the local repository.

Latinobarometro: The user can access the data download portal for the LB by clicking on the link: <https://www.latinobarometro.org/lat.jsp>. There are no register or data form requirements to access the data. The user must download the files in the Stata format for 2017, 2018, and 2020. The user must uncompress and copy the files to the folders "latinobarometro2017", "latinobarometro2018", and "latinobarometro2020" as appropriate. The user can find these folders within the "raw\_data" folder mentioned before.

ACLED: To access ACLED data, registering on the portal website is mandatory. The link to register is as follows: <https://developer.acleddata.com/>. Once registered, the user will be able to access his profile in the so-called "ACLED Access Portal" where he will have at his disposal an "access key" (for instance, "Q6CrThSB-G1UUcTL9ylZ") that will allow him to access the data from the following link: <https://acleddata.com/data-export-tool/>.

For each download request, the user must provide the "access key", and the email address used for registration. Also, the user must enable the compatibility mode as a type of data export. The user must request two downloads and therefore obtain two data files. The first corresponds to the period between 01/01/2018 and 31/12/2020, while the second corresponds to the period between 01/01/2021 and 31/12/202. Finally, the user must copy the data files to the "acled" folder inside the "raw\_data" folder.

EMDAT: To access the data, the user must register on the EMDAT portal website. The link to register is as follows: <https://public.emdat.be/>. After that, the user must access the "EM-DAT Query Tool". Once in there, the user must download data for the 2018-2022 period and limit the search to natural disasters (this excludes technological and complex disasters). The downloaded file will have a similar name to "emdat\_public\_2023\_07\_11\_query\_uid-[code].xlsx". Before copying it to the “emdat” folder within the "raw\_data" one, the user must change the downloaded file's name to "emdat\_public\_2023\_06\_18\_query\_uid-g0uFg8.xlsx". It is the only case where the user must change a data file’s name.

WVS: The user must complete a request form on the WVS portal website to access the data. The link to the form is as follows: <https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>. The user must download a file called "WVS TimeSeries 1981 2022 Stata v4 0.zip", in the "Longitudinal Data Files" section at the end of the web portal. After that, the user must uncompress and copy the data files to the "wvs6\_7" folder within the "raw\_data" folder.

EVS: The user must complete an online registration form at GESIS (Leibniz Institute for the Social Sciences). Here you can find the link to the form: <https://login.gesis.org/realms/gesis/login-actions/registration?client_id=gesis-gws-client&tab_id=zKStYMozrOQ>. Once you log into the GESIS portal you must access the EVS2017 Integrated Dataset (ZA7500) at <https://europeanvaluesstudy.eu/methodology-data-documentation/survey-2017/full-release-evs2017/documentation-survey-2017/>. The user must download the “ZA7500\_v5-0-0.dta.zip” file, it contains a Stata .dta file with the microdata. After uncompressing the data file, the user must copy the data file to the “evs” folder within the “raw\_data” folder.

To summarize these steps, Table A1 lists the files the user must download and copy to the "raw\_data" folder. Five sources and 21 files are involved in this manual data-downloading process. Finally, it is worth noting that three databases are already "included" in the SSGD v2.0 repository (as shown in Table 1); these databases are the following: CIVICUS, EIU, and EQOSOGI. We have included these data sets by default since it is impossible to download the data from a web portal. Instead, all the information within these sources comes from reports. Table A2 lists the databases included by default in the SSGD v2.0 repository.

**Table A1**: List of data files to be manually downloaded and copied to the local repository

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Access** | **Files** | **Folder** |
| ASB | <https://www.asianbarometer.org/data?page=d10> | W4\_v15\_merged20181211\_release.dta | asianbarometer4 |
| ABS Wave 5 Philippines\_Core\_merged\_20201223\_release.dta | asianbarometer5 |
| ABS\_V\_Mongolia\_merged\_core\_20201217\_release.dta |
| W5\_Australia\_merged\_core\_20210803\_release.dta |
| W5\_India\_merged\_core\_20220905\_released.dta |
| W5\_Indonesia\_merged\_core\_20220905\_released.dta |
| W5\_Japan\_merged\_core\_20220905\_released.dta |
| W5\_Korea\_merged\_core\_20210823\_released.dta |
| W5\_Malaysia\_coreQmerged\_20210819\_release.dta |
| W5\_Myanmar\_CoreQ\_20220905\_released.dta |
| W5\_Taiwan\_coreQrelease\_20190805.dta |
| W5\_Thailand\_merged\_core\_20210805\_release.dta |
| W5\_Vietnam\_merged\_core\_20201215\_release.dta |
|  |  | Latinobarometro2017Esp\_v20180117.dta | latinobarometro2017 |
| LB | <https://www.latinobarometro.org/latContents.jsp> | Latinobarometro\_2018\_Esp\_Spss\_v20190303.sav | latinobarometro2018 |
|  |  | Latinobarometro\_2020\_Esp\_Stata\_v1\_0.dta | latinobarometro2020 |
| ACLED | <https://acleddata.com/data-export-tool/> | 2018-01-01-2020-12-31.csv | acled |
| 2021-01-01-2022-12-31.csv |
| WVS | <https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp> | WVS\_TimeSeries\_4\_0.dta | wvs6\_7 |
| EVS | <https://europeanvaluesstudy.eu/methodology-data-documentation/survey-2017/full-release-evs2017/documentation-survey-2017/> | ZA7500\_v5-0-0.dta | evs |

**Table A2**: List of databases included by default in the SSGD v2.0 repository

|  |  |  |
| --- | --- | --- |
| Source | Account | Access |
| CIVICUS | People Power Under Attack 2022 | <https://civicus.contentfiles.net/media/assets/file/2022GlobalFindingsEmbargoed16March.pdf> |
| EIU | Democracy Index Reports 2015-2022 | <https://www.eiu.com/n/campaigns/democracy-index-2022/?utm_source=google&utm_medium=paid-search&utm_campaign=democracy-index-2022&gclid=EAIaIQobChMIhNOPlc-HgAMVgUBIAB3OTwxmEAAYASAAEgJjn_D_BwE> |
| EQOSOGI | Equality of Opportunity for Sexual and Gender Minorities | <https://www.worldbank.org/en/publication/equality-of-opportunity-for-sexual-and-gender-minorities> |

aIn the case of EIU, we only provide the URL for the 2022 report.

1. Contact: Paola Ballon ([pballon@worldbank.org](mailto:pballon@worldbank.org)), Omar Alburqueque ([oalburquequechav@worldbank.org](mailto:oalburquequechav@worldbank.org)). [↑](#footnote-ref-1)
2. It would not be possible to upload all the databases to the repository since it would imply a violation of the use-of-data policies. [↑](#footnote-ref-2)