# Workflow for preparing the SSP database V3

# Workflow for transforming UN Population data for Revision 2022

## F01 files

### Description

The F01 file, contains population data from 1950 through 2100, with estimates through 2021 and projections from 2022 forward. The estimates are in a single worksheet and the projections cover 10 variants, including the ‘Medium’ variant (MED). Each worksheet has 54 data columns—including population by gender (but not by age cohort), population density, measure of fertility and mortality, and migration.

Step 1: Extract the data from the file and put in a new worksheet. The purpose of this exercise is to make the data compatible for reading with GAMS.[[1]](#footnote-1) The Python code ‘preProcessF01.py’ performs this step. The program only extracts columns ‘E,K,M,N,O and BL’ of the worksheets, i.e., the UN location code, the year, and the data for population by gender and net migration. It renames the column for the UN location code to ‘UNCode’.[[2]](#footnote-2) The Python code also relabels all of the worksheets, where the names of the projections are simplified. This step produces one Excel workbook with the name “F01xlsx. N.B. The file also contains the data for the UN’s aggregate regions, such as the world (‘900’). It has two worksheets. The first, ‘Est’, contains the UN population estimates (currently from 1950-2021). Columns A and B contain respectively the UNCode and the year. Columns C:F contain the population data by gender and migration. The second worksheet, ‘Proj’, contains the combined projections, with an additional column for the projection code (in Column A and the other columns are displaced by one). (Some rows are dropped from the original Excel sheet—these are so-called separator rows and identified with the UN Codes 1802, 1803, 1828 and 1840.)

Step 2: Extract the data from the intermediate Excel workbook from Step 1, and save as a GDX file. This is done with the GAMS file called ‘readF01.gms’. It reads all the relevant labels from the GAMS file “UNHeader.gms”. It reads each worksheet. It merges the estimates and projections into a single parameter (F01Data) and the migration data, also merged, is saved in F01MigrData. The resulting merged data is saved in a GDX file called “F01\_Data.gdx”.

Step 3: Convert the file to use standard labels for countries and save the data. All the data is saved including the regional aggregations. This step is done with the GAMS file names ‘ConvF01.gms’. It reads the data from Step 2, converts the labels for country, and saves the data in a GDX file called “F01\_Data\_GTAP.gdx”. N.B. Currently the migration data is not saved in the final GDX file.

## F02 files

### Description

The F02 files, of which there are three (MALE, FEML, BOTH), contains population data from 1950 through 2100, with estimates through 2021 and projections from 2022 forward. The data is split by age cohorts of 5-years with the last one being 100 years and above. The estimates are in a single worksheet and the projections cover 10 variants, including the ‘Medium’ variant (MED).

Step 1: Extract the data from the three files and put in a new worksheet. The purpose of this exercise is to make the data compatible for reading with GAMS.[[3]](#footnote-3) The Python code ‘preProcessF02.py’ performs this step for all three workbooks. It only saves columns ‘E:K:AF’ of the workbooks, i.e., the UN location code, the year, and the data for the 21 cohorts. It renames the column for the UN location code to ‘UNCode’.[[4]](#footnote-4) The Python code also relabels all of the worksheets, where the names of the projections are simplified. This step produces three Excel workbooks with the name “F02\_[gender].xlsx. N.B. The files also contain the data for the UN’s aggregate regions, such as the world (‘900’). (Some rows are dropped from the original Excel sheet—these are so-called separator rows and identified with the UN Codes 1802, 1803, 1828 and 1840.)

Step 2: Extract the data from the three intermediate Excel workbooks from Step 1, merge and save as a GDX file. This is done with the GAMS file called ‘readF02.gms’. It reads all the relevant labels from the GAMS file “UNHeader.gms”. It reads each worksheet from each file looping over gender and the worksheet names. It calls the GAMS file ‘F02\_Data\_WKS.gms’ to read each individual worksheet which is saved temporarily in a GDX file. For each loop, the extracted worksheet data is read in from the temporary GDX file and merged with the full dataset. The resulting merged data is saved in a GDX file called “F02\_Data.gdx”.

Step 3: Convert the file to use standard labels for countries and cohort and save the data with the exception of the regional aggregates, i.e., save only the country data. This step is done with the GAMS file names ‘ConvF02.gms’. It reads the data from Step 2, converts the labels for country and cohort, and saves the data in a GDX file called “F02\_Data\_GTAP.gdx”.

## Comparison of world totals

The world total population is available in both F01 and F02, either as an aggregation of all countries or using the UN’s own aggregate world region (900). The file CompProj.gms compares the world total in all four ways. (N.B. The converted F02 GDX file does not contain the UN’s world total. It is instead extracted from the intermediate GDX file where the location codes still use the UN’s label, i.e., 900). The world totals match in three of the four cases. They do not match in F02 when aggregating the countries. The discrepancy is due to the fact that F02 does not include detailed data for the Holy See, which is in F01.

## Preparing UN data to be merged with SSP data

The F02 data is read in and prepared to be merged with the SSP data from WiC. The program is called ‘MakeSSP.gms’. It reads F02 from “F02\_Data\_GTAP.gdx”, with is by gender (including “BOTH”), UN scenario, country, and 5-year age cohorts. The WiC data has several additional dimensions—SSP and education levels of which there are 6. The merged data will be by scenario, country, gender, aggregate age groups and year. There are six scenarios—the 5 SSPs and the medium variant of the UN projections labeled “UN2022”. The aggregate age groups are “0-14”, “15-64”, “65UP”, plus a total. The UN data has full annual coverage from 1950 through 2100. The WiC data will have the UN data from 1950-2019, and the WiC data per SSP from 2020-2100 (which has been annualized using cubic splining). N.B. The WiC data is in a different folder, currently “V:/ClimateChange/SSP3/”. The original WiC data does not have the gender label “BOTH”, so this is also calculated. The merged data is saved in “SSP\_UN.gdx” and “SSP\_UN.csv”.

**Table 1: Summary workflow for UN 2022 Revision database**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WPP2022\_GEN\_F01\_DEMOGRAPHIC\_INDICATORS.xlsx** | è | **preProcessF01.py** | è | **F01.xlsx** |
|  |  |  |  |  |
| **F01.xlsx** | è | **readF01.gms** | è | **F01\_Data.gdx** |
|  |  |  |  |  |
| **F01\_Data.gdx** | è | **convF01.gms** | è | **F01\_Data\_GTAP.gdx** |
|  |  |  |  |  |
| **WPP2022\_POP\_F02\_1\_POPULATION\_5-YEAR\_AGE\_GROUPS\_BOTH\_SEXES.xlsx WPP2022\_POP\_F02\_3\_POPULATION\_5-YEAR\_AGE\_GROUPS\_FEMALE.xlsx WPP2022\_POP\_F02\_2\_POPULATION\_5-YEAR\_AGE\_GROUPS\_MALE.xlsx** | è | **preProcessF02.py** | è | **F02\_BOTH.xlsx F02\_FEML.xlsx F02\_MALE.xlsx** |
|  |  |  |  |  |
| **F02\_BOTH.xlsx F02\_FEML.xlsx F02\_MALE.xlsx** | è | **readF02.gms** | è | **F02\_Data.gdx** |
|  |  |  |  |  |
| **F02\_Data.gdx** | è | **convF02.gms** | è | **F02\_Data\_GTAP.gdx** |
|  |  |  |  |  |
| **F02\_Data\_GTAP.gdx V:/ClimateChange/SSP3/SSPV3.gdx** | è | **makeSSP.gms** | è | **SSP\_UN.gdx SSP\_UN.csv** |

# Workflow for transforming SSP data

This part describes the downloading and transformation of the SSP V3.0 data. The data is downloaded directly from the SSP [website](https://data.ece.iiasa.ac.at/ssp/#/workspaces) and various transformations are done to prepare a consolidated SSP database. The current database has data for 192 countries—corresponding to the number of countries in the OECD SSP projections file, with population broken out by gender and three broad age groups.

## Downloading the SSP database V3.0 and preliminary processing

The updated SSP database, V 3.0, was posted to the IIASA Data Explorer [website](https://data.ece.iiasa.ac.at/ssp/#/workspaces) at the end of January 2024. It contains the following:

* The WiC population projections for the period 2020-2100 in 5-year time steps with no data prior to 2020. The coverage is for 200 countries (the vast majority of the world’s population), broken out by gender, 5-year age cohorts (plus 100 and above) and by 6 education levels. The data is in millions.
* The OECD GDP projections for the period 2020-2100 in 5-year time steps. The data file also includes historical data between 1980 and 2020 (where available), also in 5-year time steps. The coverage is for 192 countries. The data is in $2017PPP billion.
* The IIASA GDP projections for the period 2025-2100 in 5-year time steps with no data prior to 2025. The coverage is for 170 countries. The data is in $2017PPP billion.

The OECD and IIASA GDP projections were downloaded as two individual ZIP files. The website struggled to download the WiC database, which was then downloaded by age cohort (WIC0\_4.zip, WIC5\_9.zip, …, WIC95-99.zip, WIC100UP.zip). All downloaded files are stored in the “Downloads” folder.

A Python script pre-processes the downloaded data (preProcess.py). It uses an input file that contains the names of the zip files to decompress and process (zipFiles.txt). It unzips each zip file and extracts any file that is not named ‘readme.txt’. In principle, each zip file contains a single CSV file that is extracted.

For all the CSV files, the script removes the last line of the CSV file, which is the IIASA copyright. There is no other processing for the OECD and IIASA GDP projections, which are stored in the same folder with the letter ‘x’ appended to the filename (e.g., OECD.csv becomes OECDx.csv). The WiC files are merged into a single CSV file called “WiC\_All.csv”. The header line of all the WiC files is discarded, except for the first file.

## Annualization of the SSP downloaded data

The downloaded SSP projections are in 5-year time steps. This step annualizes the projections (and the historical data in the OECD file) using cubic splining. This is done using the Python codes ‘SplineGDPV3.py’ and ‘SplinePopV3.py’, for the GDP and population files respectively. In addition to the annualization, the code: (1) drops the ‘Model’, ‘Variable’ and ‘Unit’ columns; (2) replaces the IIASA-based country names with the 3-letter ISO-code; and (3) in the case of the population projections, the WiC based labels for education, age cohorts and gender are replaced with a different set of labels. The mappings are provided in the code and in Annex A. The annualized data is output as a CSV file with the suffix “\_int”.

## Consolidation of the SSP files

The third step consolidates the three SSP contributions into a single file—this is done in a GAMS file named ‘ReadCSV.gms’. Two invocations of ‘Connect:CSVReader’ are used to read the two GDP projections. They are merged into a single parameter with four dimensions: models (mods), SSP, country (c) and time (t). In addition, the historical GDP trends are also stored—this parameter is indexed by model, though at this moment, there is no historical data associated with the IIASA GDP submission. The WiC population data is also read using ‘Connect:CSVReader’, and it is saved untouched, with the GDP data to the consolidated GDX file (SSPV3.gdx).

## OECD Supplemental data

The OECD has provided its projections on an annual basis—both the historical series as well as the GDP projections. The file contains total GDP (in $2017PPP billion) and GDP per capita (in $2017PPP). The OECD data for the period 2020-2030 reflects official data, when available, as well as integrating the medium-term forecasts through 2027 based on OECD and IMF reports from the Fall of 2023. This data is processed in two steps. The first step uses Python code named “preProcessOECD.py”. This reads the OECD submission (“Data for DvdM.xlsx”) and: (1) drops the model and unit columns; (2) truncates the scenario labels to the first four characters (3) pivots the data with the variables as columns; and (4) saves the data in a new Excel workbook (“OECDAnnual.xlsx”). The second step converts the Excel file from the first step into a GDX file (“OECDXLS.gdx”).

## Consolidated SSP database

The Center has downloaded and processed this data and produced a complementary SSP database, which includes the WiC population projections, the Medium Variant for population from the UN Population Division’s 2022 Revision, and annual data from the OECD projections—both historical (when available) as well as for the period 2020-2100. N.B. The OECD data for the period 2020-2030 reflects official data, when available, as well as integrating the medium-term forecasts through 2027 based on OECD and IMF reports from the Fall of 2023. This consolidation process is done using the GAMS code “MakeSSP02Feb2024.gms”.

The consolidation uses the GDP data from the OECD’s projection—using the annual data (i.e., file “OECDXLS.gdx”). The population data—historical and projections—merge the UN’s Medium Variant with WiC’s 5 SSP projections. The age cohorts are collapsed into 3 (0-14, 15-64 and 65 and above), and the education levels are dropped from the WiC projections. The set of population projections is consolidated and identified with ‘Scen’, of which there are 6 (UN, SSP1-SSP5). History for the WiC scenarios is taken from the UN from 1950 through 2019. There are minor discrepancies between WiC and the UN for the year 2020. The GDP and population data are provided in full for OECD’s 192 countries and the period 1980 through 2100. (The dropped data are available upon request.) The total GDP data from the OECD is converted to $2017PPP million to be consistent with the population scale. The population data from the UN is converted to million as well (from thousands).

The output is provided in four formats: GDX, CSV, Excel and HAR.[[5]](#footnote-5) The CSV and HAR files are split into 2—one for GDP and one for population.

The ‘official’ data in this database includes:

* The WiC data for years 2020-2100 in five year increments (albeit aggregated by age group and totaled over all education levels). The annualized data comes from the interpolation process. Data prior to 2020 is sourced from the UN 2022 Revision, which is official.
* All of the UN data is official and unadulterated, with the exception of aggregation over age groups.
* The OECD data is official between 1980 and 2100 in 5-year increments. The annual data is provided as a courtesy and is not considered official—though most of the historical data is sourced from various official databases.

**Table 2: Summary workflow for SSP database**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **OECD.csv (OECD.zip) IIASA.csv (IIASA.zip) WiC0\_4.csv (WiC0\_4.zip), WiC5\_9.csv (WiC5\_9.zip), …, WiC95\_99.csv (WiC95\_99.zip), WiC100UP.csv (WiC100UP.zip)** | è | **preProcess.py** | è | **OECDx.csv IIASAx.csv WiC\_All.csv** |
|  |  |  |  |  |
| **Downloads/OECDx.csv Downloads/IIASAx.csv** | è | **SplineGDPV3.py** | è | **OECDx\_int.csv IIASAx\_int.csv** |
|  |  |  |  |  |
| **Downloads/WiC\_All\_int.csv** | è | **SplinePopV3.py** | è | **WiC\_All\_int.csv** |
|  |  |  |  |  |
| **OECDx\_int.csv IIASAx\_int.csv WiC\_All\_int.csv** | è | **readCSV.gms** | è | **SSPV3.gdx** |
|  |  |  |  |  |
| **Data for DvdM.xlsx** | è | **preProcessOECD.py** | è | **OECDAnnual.xlsx** |
|  |  |  |  |  |
| **OECDAnnual.xlsx** | è | **readOECDXLSX.gms** | è | **OECDXLS.gdx** |
|  |  |  |  |  |
| **OECDXLS.gdx %UNFolder%/F02\_Data\_GTAP.gdx SSPV3.gdx** | è | **MakeSSP02Feb2024.gms** | è | **SSP02FEB.gdx SSP02FEBPop.csv SSP02FEBGDP.csv** |
|  |  |  |  |  |
| **SSP02FEBPop.csv SSP02FEBGDP.csv** | è | **SSPGDP.tab(.cmf,.inp)  SSPPOP.tab(.cmf,.inp)** | è | **SSP02FEBPop.HAR SSP02FEBGDP.HAR** |

# Additional processes

1. Creation of a GTAP-aggregation. The GAMS file “SSPGTAP.gms” creates a GTAP-compatible version of the consolidated SSP database. The key input file is “SSP02Feb.gdx”, which contains the GDP and population data for the 5 SSPs (from the OECD and WiC respectively), as well as the Medium Variant of the UN’s 2022 Revision for population. The data is extracted and aggregated using a mapping from countries to GTAP regions (mapGTAP). This mapping is defined for all 237 countries, though only 192 are in the database. The aggregate data is saved as a GDX file (“SSP\_GTAP02FEB.gdx”) and a CSV file, which only includes the WiC population data and total GDP. The resulting file has no data for Afghanistan (AFG), Palestine (PSE), Syria (SYR) and Venezuela (VEN), as well as two composite regions—rest of North America (XNA) and rest of the world (XTW). N.B. Some of the other composite regions may have less than complete country coverage.

**Table 3: Summary workflow for SSP GTAP-conformable database**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SSP02FEB.gdx** | è | **SSPGTAP.gms** | è | **SSP\_GTAP02FEB.gdx SSP\_GTAP02Feb.csv** |

1. The file “ReadGDX.gms” is a template file for reading the full SSP database (“SSPV3.gdx”)—which has the annualized OECD projection—not the annual OECD projection. It produces a CSV file that has the OECD GDP projections (converted to $2017PPP million) and the total population from WiC.
2. The file “CompSSP.gms” compares the SSP projections between the original V0.9 and the new V3.0. It uses the 2020 levels from V3.0 as the reference point and applies the country level projection trends from the two SSP databases. The results are saved as a GDX and CSV file—and the latter is loaded as a Pivot table into “SSPComp.xlsx”

**Table 4: Summary workflow to compare V0.9 and V3.0 of the SSP projections**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SSPV3.gdx V:/ClimateChange/AR5/SSP/sspScenV9/sspScenV9.gdx** | è | **CompSSP.gms** | è | **CompSSP.gdx SSPComp.csv** |

# Annex A: Dimensions

The UN 2022 Revision has data for 237 countries—as well as a number of regional aggregations including the world.[[6]](#footnote-6) The WiC population projections cover 200 of the 237—the missing countries are identified at the end of Table A1. The OECD projections cover 192 of the 200—the missing countries are Afghanistan (AFG), Curaçao (CUW), Guadeloupe (GLP), Martinique (MTQ), Palestine (PSE), Réunion (REU), Syria (SYR) and Venezuela (VEN). The IIASA GDP projections cover 170 countries of the 200—the missing countries are Aruba (ABW), Afghanistan (AFG), Cuba (CUB), Curaçao (CUW), Eritrea (ERI), Western Sahara (ESH), Martinique (MTQ), Micronesia (FSM), French Guiana (GUF), Guadeloupe (GLP), Guam (GUM), Guyana (GUY), Kiribati (KIR), Libya (LBY), Mayotte (MYT), New Caledonia (NCL), Papua New Guinea (PNG), Porto Rico (PRI), North Korea (PRK), French Polynesia (PYF), Réunion (REU), Solomon Islands (SLB), Somalia (SOM), South Sudan (SSD), Timor Leste (TLS), Tongo (TON), Venezuela (VEN), U.S. Virgin Islands (VIR), Vanuatu (VUT) and Samoa (WSM). N.B. The IIASA GDP projections include Syria (SYR) and Palestine (PSE), which are not part of the OECD projections.

**Table A1: List of UN countries and availability in the SSP databases**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Name | UNCode | ISO-3 | WiC | OECD | IIASA |
| 1 | Aruba | 533 | ABW |  |  | NA |
| 2 | Afghanistan | 4 | AFG |  | NA | NA |
| 3 | Angola | 24 | AGO |  |  |  |
| 5 | Albania | 8 | ALB |  |  |  |
| 7 | United Arab Emirates | 784 | ARE |  |  |  |
| 8 | Argentina | 32 | ARG |  |  |  |
| 9 | Armenia | 51 | ARM |  |  |  |
| 11 | Antigua and Barbuda | 28 | ATG |  |  |  |
| 12 | Australia | 36 | AUS |  |  |  |
| 13 | Austria | 40 | AUT |  |  |  |
| 14 | Azerbaijan | 31 | AZE |  |  |  |
| 15 | Burundi | 108 | BDI |  |  |  |
| 16 | Belgium | 56 | BEL |  |  |  |
| 17 | Benin | 204 | BEN |  |  |  |
| 19 | Burkina Faso | 854 | BFA |  |  |  |
| 20 | Bangladesh | 50 | BGD |  |  |  |
| 21 | Bulgaria | 100 | BGR |  |  |  |
| 22 | Bahrain | 48 | BHR |  |  |  |
| 23 | Bahamas | 44 | BHS |  |  |  |
| 24 | Bosnia and Herzegovina | 70 | BIH |  |  |  |
| 26 | Belarus | 112 | BLR |  |  |  |
| 27 | Belize | 84 | BLZ |  |  |  |
| 29 | Bolivia (Plurinational State of) | 68 | BOL |  |  |  |
| 30 | Brazil | 76 | BRA |  |  |  |
| 31 | Barbados | 52 | BRB |  |  |  |
| 32 | Brunei Darussalam | 96 | BRN |  |  |  |
| 33 | Bhutan | 64 | BTN |  |  |  |
| 34 | Botswana | 72 | BWA |  |  |  |
| 35 | Central African Republic | 140 | CAF |  |  |  |
| 36 | Canada | 124 | CAN |  |  |  |
| 37 | Switzerland | 756 | CHE |  |  |  |
| 38 | Chile | 152 | CHL |  |  |  |
| 39 | China | 156 | CHN |  |  |  |
| 40 | Côte d'Ivoire | 384 | CIV |  |  |  |
| 41 | Cameroon | 120 | CMR |  |  |  |
| 42 | Democratic Republic of the Congo | 180 | COD |  |  |  |
| 43 | Congo | 178 | COG |  |  |  |
| 45 | Colombia | 170 | COL |  |  |  |
| 46 | Comoros | 174 | COM |  |  |  |
| 47 | Cabo Verde | 132 | CPV |  |  |  |
| 48 | Costa Rica | 188 | CRI |  |  |  |
| 49 | Cuba | 192 | CUB |  |  | NA |
| 50 | Curaçao | 531 | CUW |  | NA | NA |
| 52 | Cyprus | 196 | CYP |  |  |  |
| 53 | Czechia | 203 | CZE |  |  |  |
| 54 | Germany | 276 | DEU |  |  |  |
| 55 | Djibouti | 262 | DJI |  |  |  |
| 57 | Denmark | 208 | DNK |  |  |  |
| 58 | Dominican Republic | 214 | DOM |  |  |  |
| 59 | Algeria | 12 | DZA |  |  |  |
| 60 | Ecuador | 218 | ECU |  |  |  |
| 61 | Egypt | 818 | EGY |  |  |  |
| 62 | Eritrea | 232 | ERI |  |  | NA |
| 63 | Western Sahara | 732 | ESH |  |  | NA |
| 64 | Spain | 724 | ESP |  |  |  |
| 65 | Estonia | 233 | EST |  |  |  |
| 66 | Ethiopia | 231 | ETH |  |  |  |
| 67 | Finland | 246 | FIN |  |  |  |
| 68 | Fiji | 242 | FJI |  |  |  |
| 70 | France | 250 | FRA |  |  |  |
| 72 | Micronesia (Fed. States of) | 583 | FSM |  |  | NA |
| 73 | Gabon | 266 | GAB |  |  |  |
| 74 | United Kingdom | 826 | GBR |  |  |  |
| 75 | Georgia | 268 | GEO |  |  |  |
| 77 | Ghana | 288 | GHA |  |  |  |
| 79 | Guinea | 324 | GIN |  |  |  |
| 80 | Guadeloupe | 312 | GLP |  | NA | NA |
| 81 | Gambia | 270 | GMB |  |  |  |
| 82 | Guinea-Bissau | 624 | GNB |  |  |  |
| 83 | Equatorial Guinea | 226 | GNQ |  |  |  |
| 84 | Greece | 300 | GRC |  |  |  |
| 85 | Grenada | 308 | GRD |  |  |  |
| 87 | Guatemala | 320 | GTM |  |  |  |
| 88 | French Guiana | 254 | GUF |  |  | NA |
| 89 | Guam | 316 | GUM |  |  | NA |
| 90 | Guyana | 328 | GUY |  |  | NA |
| 91 | China, Hong Kong SAR | 344 | HKG |  |  |  |
| 92 | Honduras | 340 | HND |  |  |  |
| 93 | Croatia | 191 | HRV |  |  |  |
| 94 | Haiti | 332 | HTI |  |  |  |
| 95 | Hungary | 348 | HUN |  |  |  |
| 96 | Indonesia | 360 | IDN |  |  |  |
| 98 | India | 356 | IND |  |  |  |
| 99 | Ireland | 372 | IRL |  |  |  |
| 100 | Iran (Islamic Republic of) | 364 | IRN |  |  |  |
| 101 | Iraq | 368 | IRQ |  |  |  |
| 102 | Iceland | 352 | ISL |  |  |  |
| 103 | Israel | 376 | ISR |  |  |  |
| 104 | Italy | 380 | ITA |  |  |  |
| 105 | Jamaica | 388 | JAM |  |  |  |
| 107 | Jordan | 400 | JOR |  |  |  |
| 108 | Japan | 392 | JPN |  |  |  |
| 109 | Kazakhstan | 398 | KAZ |  |  |  |
| 110 | Kenya | 404 | KEN |  |  |  |
| 111 | Kyrgyzstan | 417 | KGZ |  |  |  |
| 112 | Cambodia | 116 | KHM |  |  |  |
| 113 | Kiribati | 296 | KIR |  |  | NA |
| 115 | Republic of Korea | 410 | KOR |  |  |  |
| 116 | Kuwait | 414 | KWT |  |  |  |
| 117 | Lao People's Democratic Republic | 418 | LAO |  |  |  |
| 118 | Lebanon | 422 | LBN |  |  |  |
| 119 | Liberia | 430 | LBR |  |  |  |
| 120 | Libya | 434 | LBY |  |  | NA |
| 121 | Saint Lucia | 662 | LCA |  |  |  |
| 123 | Sri Lanka | 144 | LKA |  |  |  |
| 124 | Lesotho | 426 | LSO |  |  |  |
| 125 | Lithuania | 440 | LTU |  |  |  |
| 126 | Luxembourg | 442 | LUX |  |  |  |
| 127 | Latvia | 428 | LVA |  |  |  |
| 128 | China, Macao SAR | 446 | MAC |  |  |  |
| 130 | Morocco | 504 | MAR |  |  |  |
| 132 | Republic of Moldova | 498 | MDA |  |  |  |
| 133 | Madagascar | 450 | MDG |  |  |  |
| 134 | Maldives | 462 | MDV |  |  |  |
| 135 | Mexico | 484 | MEX |  |  |  |
| 137 | North Macedonia | 807 | MKD |  |  |  |
| 138 | Mali | 466 | MLI |  |  |  |
| 139 | Malta | 470 | MLT |  |  |  |
| 140 | Myanmar | 104 | MMR |  |  |  |
| 141 | Montenegro | 499 | MNE |  |  |  |
| 142 | Mongolia | 496 | MNG |  |  |  |
| 144 | Mozambique | 508 | MOZ |  |  |  |
| 145 | Mauritania | 478 | MRT |  |  |  |
| 147 | Martinique | 474 | MTQ |  | NA | NA |
| 148 | Mauritius | 480 | MUS |  |  |  |
| 149 | Malawi | 454 | MWI |  |  |  |
| 150 | Malaysia | 458 | MYS |  |  |  |
| 151 | Mayotte | 175 | MYT |  |  | NA |
| 152 | Namibia | 516 | NAM |  |  |  |
| 153 | New Caledonia | 540 | NCL |  |  | NA |
| 154 | Niger | 562 | NER |  |  |  |
| 155 | Nigeria | 566 | NGA |  |  |  |
| 156 | Nicaragua | 558 | NIC |  |  |  |
| 158 | Netherlands | 528 | NLD |  |  |  |
| 159 | Norway | 578 | NOR |  |  |  |
| 160 | Nepal | 524 | NPL |  |  |  |
| 162 | New Zealand | 554 | NZL |  |  |  |
| 163 | Oman | 512 | OMN |  |  |  |
| 164 | Pakistan | 586 | PAK |  |  |  |
| 165 | Panama | 591 | PAN |  |  |  |
| 166 | Peru | 604 | PER |  |  |  |
| 167 | Philippines | 608 | PHL |  |  |  |
| 169 | Papua New Guinea | 598 | PNG |  |  | NA |
| 170 | Poland | 616 | POL |  |  |  |
| 171 | Puerto Rico | 630 | PRI |  |  | NA |
| 172 | Dem. People's Republic of Korea | 408 | PRK |  |  | NA |
| 173 | Portugal | 620 | PRT |  |  |  |
| 174 | Paraguay | 600 | PRY |  |  |  |
| 175 | State of Palestine | 275 | PSE |  | NA |  |
| 176 | French Polynesia | 258 | PYF |  |  | NA |
| 177 | Qatar | 634 | QAT |  |  |  |
| 178 | Réunion | 638 | REU |  | NA | NA |
| 179 | Romania | 642 | ROU |  |  |  |
| 180 | Russian Federation | 643 | RUS |  |  |  |
| 181 | Rwanda | 646 | RWA |  |  |  |
| 182 | Saudi Arabia | 682 | SAU |  |  |  |
| 183 | Sudan | 729 | SDN |  |  |  |
| 184 | Senegal | 686 | SEN |  |  |  |
| 185 | Singapore | 702 | SGP |  |  |  |
| 187 | Solomon Islands | 90 | SLB |  |  | NA |
| 188 | Sierra Leone | 694 | SLE |  |  |  |
| 189 | El Salvador | 222 | SLV |  |  |  |
| 191 | Somalia | 706 | SOM |  |  | NA |
| 193 | Serbia | 688 | SRB |  |  |  |
| 194 | South Sudan | 728 | SSD |  |  | NA |
| 195 | Sao Tome and Principe | 678 | STP |  |  |  |
| 196 | Suriname | 740 | SUR |  |  |  |
| 197 | Slovakia | 703 | SVK |  |  |  |
| 198 | Slovenia | 705 | SVN |  |  |  |
| 199 | Sweden | 752 | SWE |  |  |  |
| 200 | Eswatini | 748 | SWZ |  |  |  |
| 202 | Seychelles | 690 | SYC |  |  |  |
| 203 | Syrian Arab Republic | 760 | SYR |  | NA |  |
| 205 | Chad | 148 | TCD |  |  |  |
| 206 | Togo | 768 | TGO |  |  |  |
| 207 | Thailand | 764 | THA |  |  |  |
| 208 | Tajikistan | 762 | TJK |  |  |  |
| 210 | Turkmenistan | 795 | TKM |  |  |  |
| 211 | Timor-Leste | 626 | TLS |  |  | NA |
| 212 | Tonga | 776 | TON |  |  | NA |
| 213 | Trinidad and Tobago | 780 | TTO |  |  |  |
| 214 | Tunisia | 788 | TUN |  |  |  |
| 215 | Türkiye | 792 | TUR |  |  |  |
| 217 | Taiwan | 158 | TWN |  |  |  |
| 218 | United Republic of Tanzania | 834 | TZA |  |  |  |
| 219 | Uganda | 800 | UGA |  |  |  |
| 220 | Ukraine | 804 | UKR |  |  |  |
| 221 | Uruguay | 858 | URY |  |  |  |
| 222 | United States of America | 840 | USA |  |  |  |
| 223 | Uzbekistan | 860 | UZB |  |  |  |
| 225 | Saint Vincent and the Grenadines | 670 | VCT |  |  |  |
| 226 | Venezuela (Bolivarian Republic of) | 862 | VEN |  | NA | NA |
| 228 | United States Virgin Islands | 850 | VIR |  |  | NA |
| 229 | Viet Nam | 704 | VNM |  |  |  |
| 230 | Vanuatu | 548 | VUT |  |  | NA |
| 232 | Samoa | 882 | WSM |  |  | NA |
| 234 | Yemen | 887 | YEM |  |  |  |
| 235 | South Africa | 710 | ZAF |  |  |  |
| 236 | Zambia | 894 | ZMB |  |  |  |
| 237 | Zimbabwe | 716 | ZWE |  |  |  |
| 4 | Anguilla | 660 | AIA | NA | NA | NA |
| 6 | Andorra | 20 | AND | NA | NA | NA |
| 10 | American Samoa | 16 | ASM | NA | NA | NA |
| 18 | Bonaire, Sint Eustatius and Saba | 535 | BES | NA | NA | NA |
| 25 | Saint Barthélemy | 652 | BLM | NA | NA | NA |
| 28 | Bermuda | 60 | BMU | NA | NA | NA |
| 44 | Cook Islands | 184 | COK | NA | NA | NA |
| 51 | Cayman Islands | 136 | CYM | NA | NA | NA |
| 56 | Dominica | 212 | DMA | NA | NA | NA |
| 69 | Falkland Islands (Malvinas) | 238 | FLK | NA | NA | NA |
| 71 | Faroe Islands | 234 | FRO | NA | NA | NA |
| 76 | Guernsey | 831 | GGY | NA | NA | NA |
| 78 | Gibraltar | 292 | GIB | NA | NA | NA |
| 86 | Greenland | 304 | GRL | NA | NA | NA |
| 97 | Isle of Man | 833 | IMN | NA | NA | NA |
| 106 | Jersey | 832 | JEY | NA | NA | NA |
| 114 | Saint Kitts and Nevis | 659 | KNA | NA | NA | NA |
| 122 | Liechtenstein | 438 | LIE | NA | NA | NA |
| 129 | Saint Martin (French part) | 663 | MAF | NA | NA | NA |
| 131 | Monaco | 492 | MCO | NA | NA | NA |
| 136 | Marshall Islands | 584 | MHL | NA | NA | NA |
| 143 | Northern Mariana Islands | 580 | MNP | NA | NA | NA |
| 146 | Montserrat | 500 | MSR | NA | NA | NA |
| 157 | Niue | 570 | NIU | NA | NA | NA |
| 161 | Nauru | 520 | NRU | NA | NA | NA |
| 168 | Palau | 585 | PLW | NA | NA | NA |
| 186 | Saint Helena | 654 | SHN | NA | NA | NA |
| 190 | San Marino | 674 | SMR | NA | NA | NA |
| 192 | Saint Pierre and Miquelon | 666 | SPM | NA | NA | NA |
| 201 | Sint Maarten (Dutch part) | 534 | SXM | NA | NA | NA |
| 204 | Turks and Caicos Islands | 796 | TCA | NA | NA | NA |
| 209 | Tokelau | 772 | TKL | NA | NA | NA |
| 216 | Tuvalu | 798 | TUV | NA | NA | NA |
| 224 | Holy See | 336 | VAT | NA | NA | NA |
| 227 | British Virgin Islands | 92 | VGB | NA | NA | NA |
| 231 | Wallis and Futuna Islands | 876 | WLF | NA | NA | NA |
| 233 | Kosovo (under UNSC res. 1244) | 412 | XKX | NA | NA | NA |

**Table A2: List of scenarios**

|  |  |
| --- | --- |
| Code | Description |
| UN2022 | UN Population Division’s 2022 Revision (Medium variant only) |
| SSP1 | Sustainability ('Taking the Green Road') |
| SSP2 | Middle of the Road |
| SSP3 | Regional Rivalry ('A Rocky Road') |
| SSP4 | Inequality ('A Road Divided') |
| SSP5 | Fossil-fueled Development ('Taking the Highway') |

*Note*: The GDP projections are only available for the 5 SSPs

**Table A3: List of age cohorts**

|  |  |  |
| --- | --- | --- |
| Original | New |  |
| Code | Code | Description |
| 0-4 | P0004 | Population ages 00-04 |
| 5-9 | P0509 | Population ages 05-09 |
| 10-14 | P1014 | Population ages 10-14 |
| 15-19 | P1519 | Population ages 15-19 |
| 20-24 | P2024 | Population ages 20-24 |
| 25-29 | P2529 | Population ages 25-29 |
| 30-34 | P3034 | Population ages 30-34 |
| 35-39 | P3539 | Population ages 35-39 |
| 40-44 | P4044 | Population ages 40-44 |
| 45-49 | P4549 | Population ages 45-49 |
| 50-54 | P5054 | Population ages 50-54 |
| 55-59 | P5559 | Population ages 55-59 |
| 60-64 | P6064 | Population ages 60-64 |
| 65-69 | P6569 | Population ages 65-69 |
| 70-74 | P7074 | Population ages 70-74 |
| 75-79 | P7579 | Population ages 75-79 |
| 80-84 | P8084 | Population ages 80-84 |
| 85-89 | P8589 | Population ages 85-89 |
| 90-94 | P9094 | Population ages 90-94 |
| 95-99 | P9599 | Population ages 95-99 |
| 100+ | P100P | Population ages 100 and above |

**Table A4: List of aggregate age cohorts**

|  |  |
| --- | --- |
| Code | Description |
| PLT15 | Population aged 0 to 14 |
| P1564 | Population aged 15 to 64 |
| P65UP | Population aged 65 and up |
| PTOTL | Total population |

**Table A5: List of gender**

|  |  |
| --- | --- |
| Code | Description |
| MALE | Male |
| FEML | Female |
| BOTH | Both |

**Table A6: List of GDP variables**

|  |  |
| --- | --- |
| Code | Description |
| GDP|PPP | (Total) GDP in $2017PPP million |
| GDP\_per\_capita|PPP | GDP per capita in $2017PPP |

**Table A7: List of education levels**

|  |  |
| --- | --- |
| Code | Description |
| NONE | No Education |
| PRMX | Incomplete Primary Education |
| PRIM | Primary Education |
| SECL | Lower Secondary Education |
| SECU | Upper Secondary Education |
| TERT | Post-Secondary Education |

*Notes*: (1) The labels in the original WiC file correspond to the ‘Description’ column in the table. (2) The education levels in the WiC database have been aggregated away in the consolidated SSP file.

**Table A8: List of models**

|  |  |
| --- | --- |
| Code | Description |
| IIASA | IIASA |
| OECD | OECD |

# Annex B: Converting a CSV file to HAR

Step 1: Convert a CSV file to HAR using the utility ‘csv2har.exe’. It is easiest to prepare an input file (.INP) file with the structure of the CSV file. For the GDP file, the input file has the following information:

SSP02FebGDP.csv ! name of input file

SSP02FebGDP0.har ! name of output file

5 ! column no for values

4 ! number of index sets (columns)

1 VAR ! Column name: GDP variable

2 SCEN ! Column name: Scenario

3 ISO ! Column name: Country ISO code

4 YEAR ! Column name: Year

The population input file has the following:

SSP02Febpop.csv ! name of input file

SSP02Febpop0.har ! name of output file

6 ! column no for values

5 ! number of index sets (columns)

1 SCEN ! Column name: Scenario

2 ISO ! Column name: Country ISO code

3 GNDR ! Column name: Gender

4 CHRT ! Column name: Age cohort

5 YEAR ! Column name: Year

This step creates 2 HAR files (with the suffix 0), but the files contain non-informative set definitions.

Step 2: Convert the HAR files from the ‘csv2har’ program to have meaningful set definitions. This step requires using GEMPACK to convert the set definitions and thus requires a ‘TAB’ and ‘CMF’ file. The GDP versions of the files are listed below. The population versions are available upon request. The output file has the same filename, with the ‘0’ suffix dropped.

**File**  
 InFile *# input file har file from csv2har #*;  
 (**new**)OutFile *# output file #*;  
  
**Set**  
 VAR *# Variable name #* **read** **elements** **from** **file** InFile **header** *"SET1"*;  
 SCEN *# SSP scenarios #* **read** **elements** **from** **file** InFile **header** *"SET2"*;  
 ISO *# ISO country code #* **read** **elements** **from** **file** InFile **header** *"SET3"*;  
 YEAR *# Year #* **read** **elements** **from** **file** InFile **header** *"SET4"*;  
**Coefficient**  
 (**all**,v,VAR)(**all**,s,SCEN)(**all**,c,ISO)(**all**,t,YEAR) CSVData(v,s,c,t) *# Array extracted from CSV #*;  
**Read**  
 CSVData **from** **file** InFile **header** *"CSV"*;  
**Coefficient**  
 (**all**,v,VAR)(**all**,s,SCEN)(**all**,c,ISO)(**all**,t,YEAR) GDPData(v,s,c,t) *# Clean data #*;  
**Formula**  
 (**all**,v,VAR)(**all**,s,SCEN)(**all**,c,ISO)(**all**,t,YEAR) GDPData(v,s,c,t) = CSVData(v,s,c,t);  
**Write**  
 (**set**)VAR **to** **file** OutFile **header** *"VAR"*;  
 (**set**)SCEN **to** **file** OutFile **header** *"SCEN"*;  
 (**set**)ISO **to** **file** OutFile **header** *"ISO"*;  
 (**set**)YEAR **to** **file** OutFile **header** *"YEAR"*;  
 GDPData **to** **file** OutFile **header** *"DATA"*;

**auxiliary files** = SSPGDP ;  
**check-on-read** **elements** = **warn**;  
**cpu**=**yes** ;  
**log file** = **yes**;  
  
**File** InFile = SSP02FebGDP0.har; *! input file har file from csv2har***File** OutFile = SSP02FebGDP.har; *! output file*

1. I cannot get GAMS to read the original UN workbooks using GAMS connect. I am also no longer able to use GAMS transfer with Python, which would have merged steps 1 and 2. [↑](#footnote-ref-1)
2. The original column label is “Location code”, which is not a valid set name in GAMS. [↑](#footnote-ref-2)
3. I cannot get GAMS to read the original UN workbooks using GAMS connect. I am also no longer able to use GAMS transfer with Python, which would have merged steps 1 and 2. [↑](#footnote-ref-3)
4. The original column label is “Location code”, which is not a valid set name in GAMS. [↑](#footnote-ref-4)
5. The Excel workbook is created by reading in the two individual CSV files into respective Pivot tables. [↑](#footnote-ref-5)
6. The Holy See is included in some of the Excel files (e.g., F01), but not all (e.g., F02). The world total (‘900’) includes the Holy See in both cases. However, in the case of the F02 data, own aggregation of the 236 countries in the F02 file will lead to a discrepancy with the world total—the difference being the population total from the Holy See. [↑](#footnote-ref-6)