SWS Population Domain

Francesca Rosa
Food and Agriculture Organization
of the United Nations

Abstract

Population: this document contains information on the overall workflow to populate the new population datasets stored in the SWS. It replies to many questions as for example: why have these datasets been created? How can these datasets be updated when a new release from the data-source is available?

Keywords: Population, World Population Prospects, World Urbanization Prospect.

1. Purpose

Several statistical processes already embedded in the SWS depend on the Population. For example, the per capita Dietary Energy Supply, by definition, can be computed as long as we dispose of reliable population data.

In addition, since many Statistical processes depend on the Population, it is extremely important to ensure that all those processes point to the same Population variables in order to ensure the compliance and the comparability between different environments in the SWS.

The purpose of the recent review of the Population domain is to identify a unique source for Population data.

This paragraph contains information about the Population data already stored in the SWS in the dataset *Population*.

In particular, the dataset `Population` contains three elements:

- *Total Population* [`element-code 11`]
- *Food balance sheets population*, [`element-code 21`]
- *Weighted Total Population*, [`element-code 511`]

Apparently, these variables have not been updated anymore, and we do not dispose of broad set of metadata to rebuild the statistical processes performed to compute this figures.

2. The new UNPD dataset

It has been agreed that the only source for Population data is the **United Nations Population Departement** (UNPD) who disseminates the *World Population Prospects* (WPP) and the *World Urbanization Prospect* (WUP). The UNPD database can be directly accessed from the UNDP URL.

The elements inherited from the UNPD data-source are:

- Total population (WPP), [`element-code: 511`]

- Female population (WPP), [element-code: 513]
- Male population (WPP), [element-code: 512]
- Urban population (WUP), [element-code: 561]
- Rural population (WUP), [element-code: 551]

These variables are already published on FAOSTAT. In order to ensure the compliance in terms of classifications (between FAOSTAT and the SWS) the same element-codes have been used to store these elements in the SWS.

IMPORTANT NOTE: please note that the **new release of the World Population Prospects and the World Urbanization Prospects may not occur at the same time**. The implication of this misalignment in the release-period of the WPP and the WUP is, mainly, the possible misalignment of the aggregated data for example data referring to regions, sub-regions and income groups that might have been designed in different moments. For this reason, it is recommended to look separately at the following documents:

- World Population Prospects: The 2017 Revision - CLASSIFICATION OF COUNTRIES BY REGION, INCOME GROUP AND SUBREGION OF THE WORLD
- World Urbanization Prospects: The 2014 Revision - CLASSIFICATION OF COUNTRIES BY MAJOR AREA AND REGION OF THE WORLD AND INCOME GROUP

The World Population Prospects also contains long terms projection of the Total, Female and Male Population. The projections currently included in the SWS are under the assumptions of "Medium variant fertility" [1].

All this information is stored in the metadata in order to support the users to properly interpret data. In particular, the metadata at observation level contains information on:

- the data-source which can be either the *World Population Prospect* (WPP) or the *World Urbanization Prospect* (WUP);
- the year of the last release;
- if the figure is an estimation or a projection and, in this latter case, the method used to compute the projection, so far *Medium Variant*

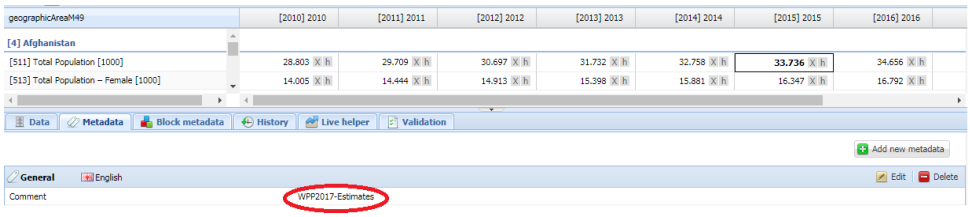


Figure 1: Metadata example

Two dataset has been created to host in the SWS the UNPD population data:

[1] Additional information is available in the UNPD document: Methodology of the United Nations Population Estimates and Projections, 2017 Revision

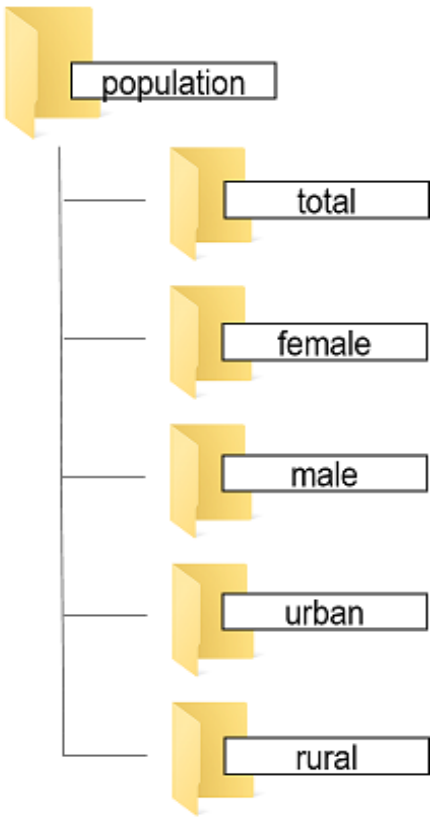
- **Source Population UNPD**, this dataset represents a clone of the data disseminated by UNPD. It basically contains all the UNPD data. The reason why it has been decided to create this dataset is that despite both UNPD and FAO support the same international classification for the geographic area, the perfect compliance is ensured only at country-level while there is no match in terms of M49 codes for regions, sub-regions and income groups. It has been decided to host, in any case, all the UNPD data, since *countries or areas listed individually are only those with 90,000 inhabitants or more in 2017; the rest are included in the aggregates but are not listed separately.*[2]
- **Population UNPD**, this detasets is the *operational* one. It contains only the M49 codes supported by the FAO. This means that all the other statistical processes embedded in the `SWS` should point to this dataset.

3. How to update new UNPD datasets

UNPD Data must be manually downloaded from the UNPD download page. A routine has been created to reshape the data in order to be hosted by the `SWS`. This routine depends on the structure of the files downloaded and can be re-used as long as the UNPD files have the same template.

Some additional operations have to be performed in order to locally store the data in a way that is suitable to be imported in the `SWS`.

1. The first step is to create, within the R work directory, the folder structure to host the UNPD data. The R routine contains the path to access to each file so it is important to reproduce exactly the same folder and sub-folder structure unless the path is properly changed.



[2] UNPD

Figure 2: Folder structure

2. As already mentioned, UNPD data may come from the *World Population Prospects* (Total population, Female and Male population) and from the *World Urbanization Prospect* (Rural and Urban population). Data coming from the *World Population Prospects* contains long terms projections which are usually stored in a separated excel sheet. The second step is to extract from the downloaded excel files the proper .csv files that can be imported into `R script` to be reshaped and saved in the `SWS`.

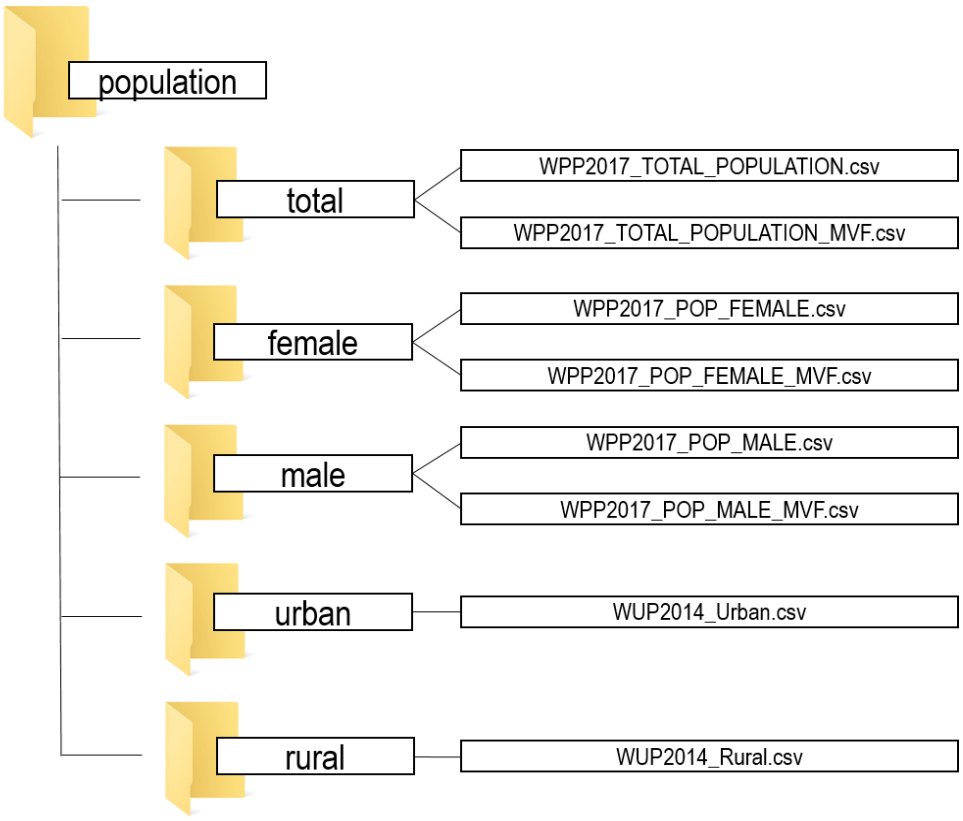


Figure 3: The files WPP2017_TOTAL_POPULATION_MVF.csv, PP2017_TOTAL_FEMALE_MVF.csv, PP2017_TOTAL_MALE_MVF.csv contains the Medium Fertility Variant projections.

3. Clean the .csv files: delete the initial colums and rename the *Area* column as reported in the following picture containing the template to be adopted.

A	B	C	D	E	F	G	H
Variant	area	Notes	Country_code	1950	1951	1952	1953
Estimates	WORLD		900	2 536 275	2 583 817	2 630 584	2 677 230
Estimates	More developed regions	a	901	814 865	824 213	834 074	844 264
Estimates	Less developed regions	b	902	1 721 410	1 759 604	1 796 510	1 832 967
Estimates	Least developed countries	c	941	195 259	199 052	202 905	206 885
Estimates	Less developed regions, excluding least developed countries	d	934	1 526 151	1 560 552	1 593 605	1 626 082
Estimates	Less developed regions, excluding China		948	1 157 197	1 179 836	1 203 955	1 229 495
Estimates	High-income countries	e	1503	672 896	680 630	688 859	697 440
Estimates	Middle-income countries	e	1517	1 734 481	1 772 394	1 808 889	1 844 730
Estimates	Upper-middle-income countries	e	1502	956 204	980 195	1 001 876	1 022 054

Figure 4: CSV structure example

4. Open a new session on the `SWS` on the dataset `Source UNPD population`. Create a new

token[3] using the only available plugin labelled **Population Token**. The token is obtained clicking on the "New debug session" button.

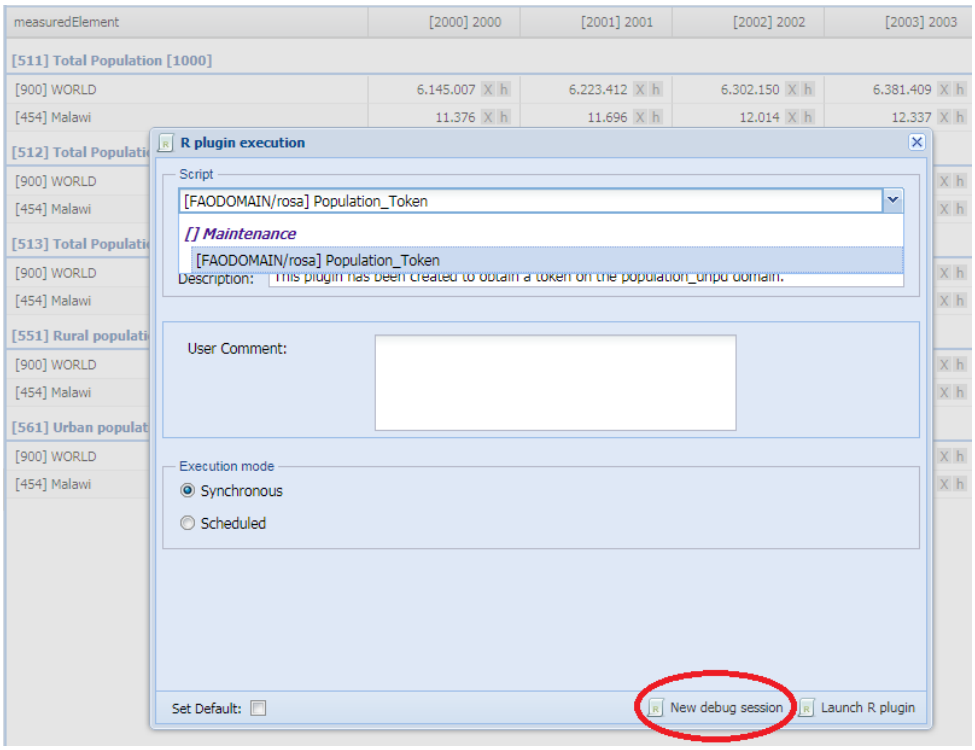


Figure 5: Create a Token from the dataset, Source UNPD population

5. Use the token[4] to perform the sub-routine stored in the Rstudio *Population project*: **convertUNPDdata.R**.

   This routine:

   - assigns the SWS elements to the different UNPD variables;
   - assigns SWS flags;
   - builds the SWS metadata file. Please specify in the routine the version of both the WPP and WUP (line 1-2). For example, data currently hosted in the SWS are classified as *"WPP2017"/"WUP2014"*;
   - populates the Source UNPD population dataset with the new figures.

Once the Source UNPD Population dataset has been populated, it is possible to open a new session directly in the SWS on the **UNPD population** dataset and run the **UNPD synchronization plugin** that populates the *operational dataset* with only the M49 compliant geographic codes.

Affiliation:

Francesca Rosa
Economics and Social Statistics Division (ESS)
Economic and Social Development Department (ES)

[3]The token ensure the connection between the R routine and the SWS. It contains the cohordinates of the SWS session where the data has to to be saved.

[4]This means that the token must be copied in the sws.yml file stored in the local-directory

Food and Agriculture Organization of the United Nations (FAO)
Viale delle Terme di Caracalla 00153 Rome, Italy
E-mail: Francesca.Rosa@fao.org
URL: https://svn.fao.org/projects/SWS/RModules/faoswsProduction/