SWS Population Domain

Charlotte Taglioni (Statistician, ESSD - Methodological Innovation Team)
Food and Agriculture Organization of the United Nations

Introduction

Population data are a key part of several statistical processes and many of them are already embedded in the Statistical Working System (SWS), e.g. the per capita Dietary Energy Supply.

In order to ensure compliance and comparability among SWS environments using Population data, it is extremely important to ensure that all SWS processes point to the same Population variables.

It has been agreed that the only source for Population data ***by country*** is the United Nations Population Department (UNPD) who disseminates the World Population Prospects (WPP) and the World Urbanization Prospect (WUP). The UNPD database can be directly accessed from the official url https://www.un.org/en/development/desa/population/index.asp.

An important topic that has been discussed in a meeting on September 4th 2019 is the dissemination of aggregates in FAOSTAT. The discussion is summarized in the document 'Note_on_aggregates_in_Population Statistics FAOSTAT.docx'.
Since the aggregates disseminates by the Population Division (UNDESA - UNPD) respect the current geopolitical asset whereas FAO disseminates aggregates according to the historical geopolitical asset, FAO only considers country specific data and then calculates aggregates independently. This decision implied also the creation of additional code discussed in the following paragraph: *Population domain additional codes.*
All the other sections are specific to the SWS domain and illustrate the step by step procedure to update Population data in the SWS.

Population domain additional codes

In the data disseminated by UNPD there might be missing figures especially for small country sex specific data. The reason relates to the decision to disseminate only countries with more than 90,000 inhabitants, as specified in the 'Note' spreadsheet of the WPP files:

'*Countries or areas listed individually are only those with 90,000 inhabitants or more in 2019; the rest are included in the aggregates but are not listed separately. The designations employed and the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory or area or its authorities, or concerning the delimitation of its frontiers or boundaries. The names and composition of geographical areas follow those presented in "Standard country or area codes for statistical use" (ST/ESA/STAT/SER.M/49/Rev.3), available at http://unstats.un.org/unsd/methods/m49/m49.htm.'*

This is relevant for FAO in order to ensure consistency among data disseminated and aggregate calculation.

In particular, nine regions have been identified for which there are missing single country data: Caribbean, Micronesia, Northern America, Northern Europe, Polynesia, South America, Southern Europe, Western Africa and Western Europe.

For these nine regions, nine additional codes have been created serving as residual variables including the countries not singularly reported. Table 1 reports the full list both with FAOSTAT and M49 standard[1] codes.

ad-hoc area name	ad-hoc area code FAOSTAT	ad-hoc area code M49
Caribbean, unspecified for population	609	029.01
Micronesia, unspecified for population	617	057.01
Northern America, unspecified for population	622	021.01
Northern Europe, unspecified for population	619	154.01
Polynesia, unspecified for population	618	061.01
South America, unspecified for population	614	005.01
Southern Europe, unspecified for population	620	039.01
Western Africa, unspecified for population	616	011.01
Western Europe, unspecified for population	621	155.01

Table 1

Another note has to be pointed out about codes used for China.
For SWS internal reasons and ensure any domain to be able to use population data, the data for China have been assigned both to codes '1248' and '156'. It is important to bear this in mind when calculating aggregates to avoid duplications.

[1] Please note these codes are not officially included in the UNSD M49 standard, nevertheless they are based on M49 classification.

Migration actors

Any migration into the SWS requires the interaction between at least three actors:

1. The **technical unit** interested in automating either some or all its analytic processes in the SWS. In the Population domain migration the technical unit is the **Social Statistics team** and it is represented by:

- **Piero Conforti**, ESS Deputy Director and team leader

- **Gurbuzer Yonca**, Statistician (ESSD)

With the support of **FAOSTAT** focal point:

- Amanda Gordon, Statistician (ESSD)

2. The **ESS - Methodological Innovation Team (ESS - SWS)** responsible for the implementation and documentation of the required statistical processes. From the ESS - SWS team the focal points for FIAS - SWS process migration are:

- **Carola Fabi**, Senior Statistician and team leader (ESSD)

- **Charlotte Taglioni**, Statistician (ESSD)

3. The **CSI - SWS** team, the primary backend and frontend maintainer of the SWS and responsible for the implementation and documentation of non-statistical processes (IT infrastructure). The CSI - SWS team have as interlocutors:

- **John Rowell**, Information Technology Officer (CSI)

- **Enrico Anello**, Full Stack Developer (CSI)

- **Matteo Terrinoni**, Lead Front End developer (CSI)

Domain structure

Despite its simple structure, the population domain has been revised several times and its process has been significantly simplified during the last revision.

The domain includes a harvester, two datasets, three datatables and a plugin:

Harvester:

- **Population UNDP source – Harvester** (*population_undp_harvester*)

Datasets:

- **Population UNPD (Source)** (*source_population_undp*): it contains the data coming from the WPP and the WUP as published by UNDESA including aggregates
- **Population UNPD** (*population_undp*): it contains the country data that will be used and disseminated in FAOSTAT

Datatables:

- **geographic_hierarchy_ebx5**: it contains the M49 hierarchy as saved in EBX5 and allows to group data by intermediate regions, sub regions, region and at global level.
- **pop_code_conversion_unpd_fao**: contains both code used by UNDESA and those used by FAO. Country codes are all compliant as they respect the M49 codes but for higher geographic level the codes might differ.
- **pop_unspecified_conversion**: it contains the additional 'Population unspecified' codes with the correspondent UNPD and FAOSTAT codes.

Plugin:

- **population_unpd2fao**: run from the Population UNPD dataset.

Update Population dataset

The first step of the procedure is to download the newest **EXCEL** release of the WPP and WUP respectively at these links:

- WPP: https://population.un.org/wpp/Download/Standard/Population/ download 'Total Population - Both Sexes', 'Total Population – Male' and 'Total Population – Female' files.
- WUP: https://population.un.org/wup/Download/ download 'WUP2018-F19-Urban_Population_Annual.xls' and 'WUP2018-F20-Rural_Population_Annual.xls' files.

A total of 5 Excel files is need as input for the harvester.

After the download of these files the Population harvester in SWS can be launched.

Once on the initial page of the SWS select the 'Harvester' button as in Figure 1.

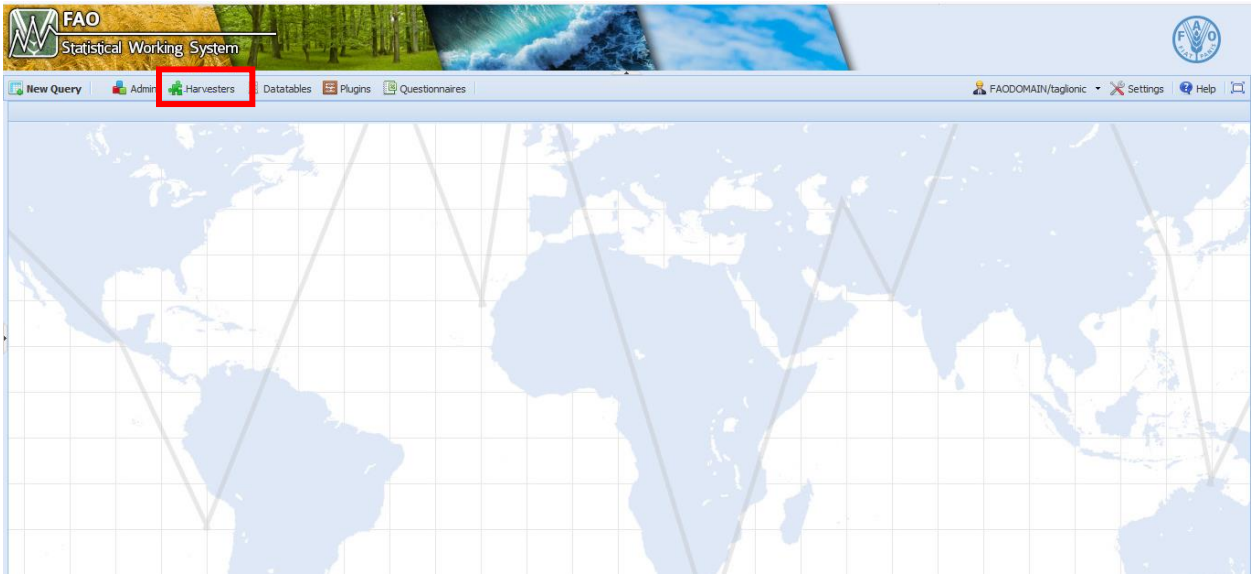

Figure 1

From the Job list select the 'Population UNDP source - Harvester' option (Figure 2)

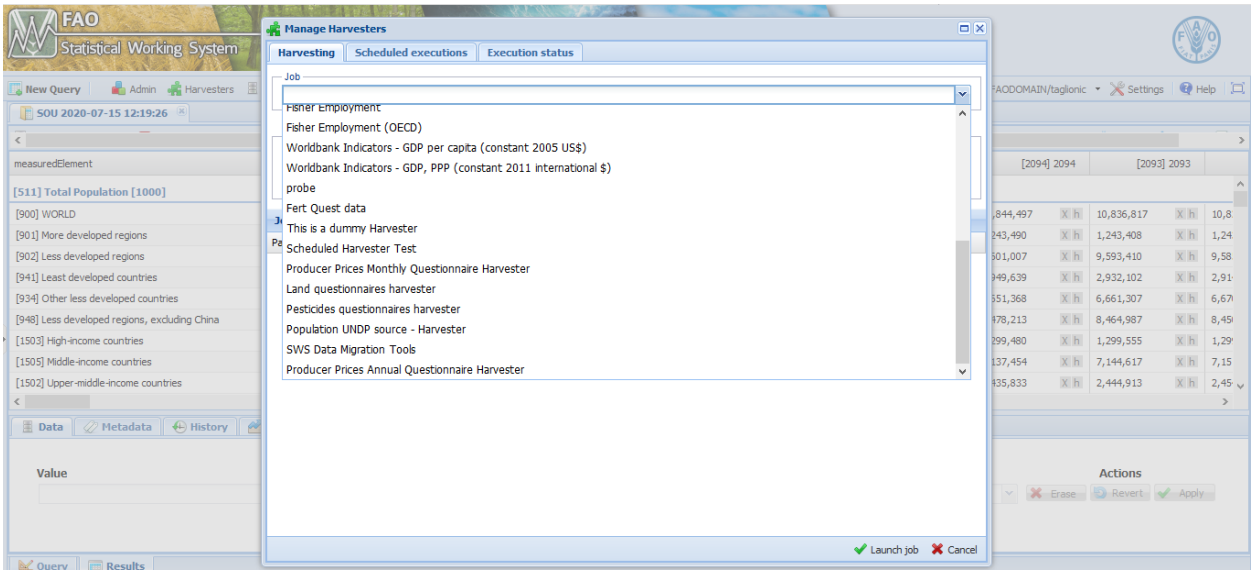

Figure 2

When the Population harvester tab is open, in the 'Job input params' section select for each parameter the correspondent Excel file downloaded and click on the button 'Launch job' (Figure 3).

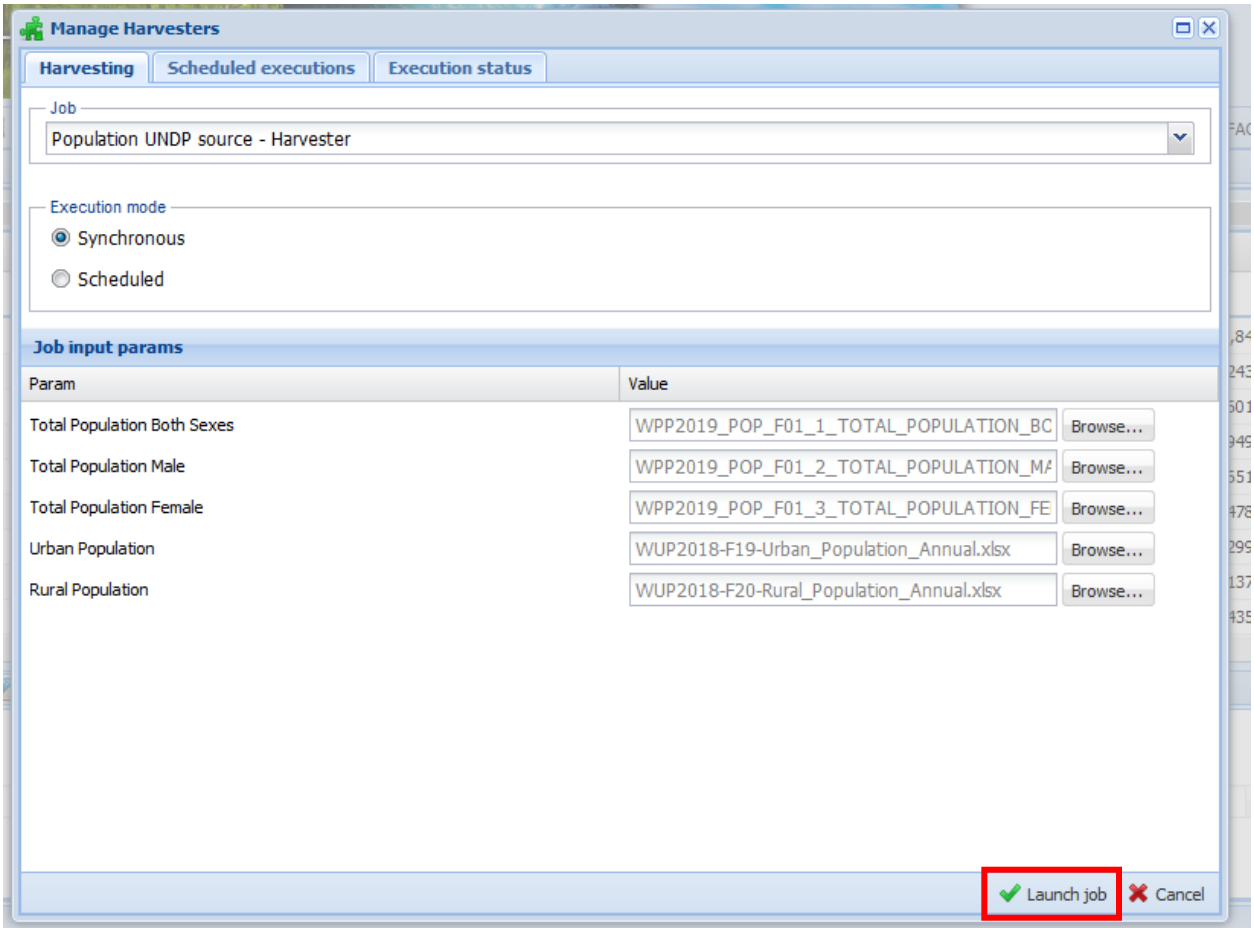

Figure 3

When the harvester process is completed, the user will receive an email confirming the success of the operation. If the email announce the failure of the procedure please contact the CSI team.

Please note there is a list of code used by the UNPD but not included in SWS that will be disregarded by the system. These codes mainly refer to SDG regions and ignoring them does not affect the overall procedure. These codes are: 1803, 1636, 1637, 1802, 1518, 1840, 1828, 1833, 1832, 1830, 1835, 1829, 917, 918. If an additional code appears in the .csv attached to the confirmation email please contact the Methodological team.

Once the email comunicates the data have been saved in the dataset please check the Population UNPD (source) dataset to ensure the procedure has worked well. From the initial page click the 'New Query' button and then select the 'Population' domain and the 'Populaiton UNPD (source)' dataset (Figure 4).

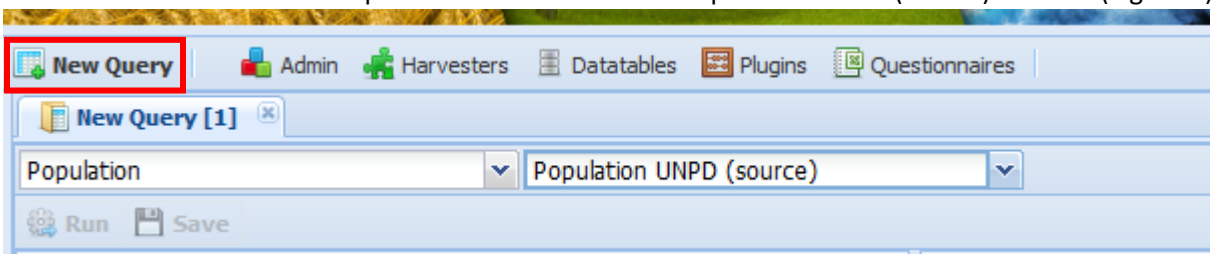

Figure 4

When the dimension display select either a sample or the whole dataset and then click on the 'Run' button to visualize the selected session (Figure 5).

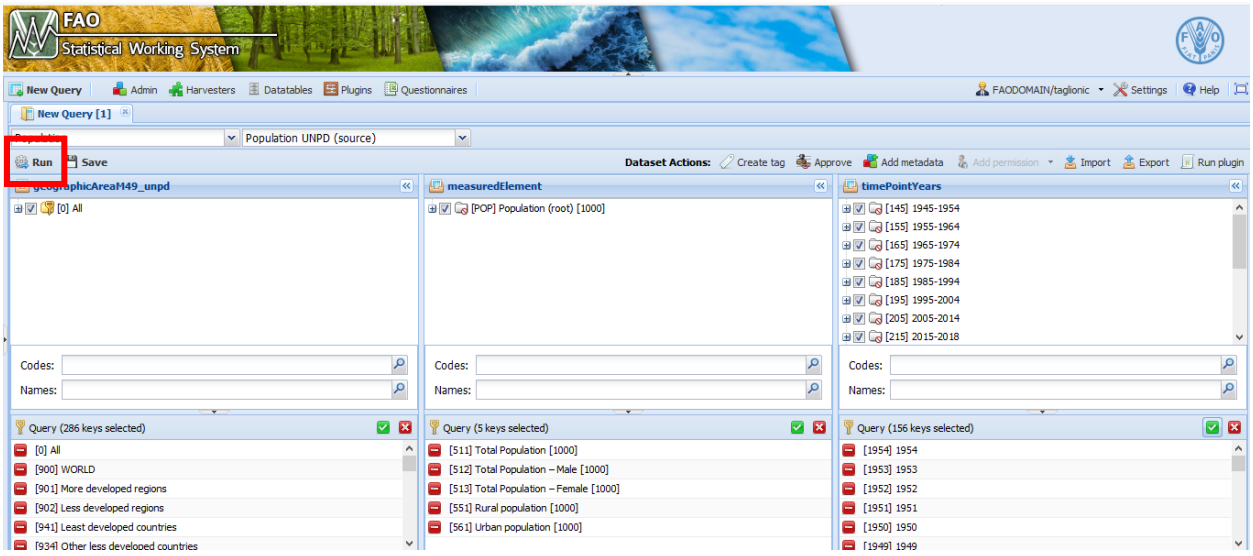

Figure 5

Once data have been checked, the plugin to populate the main dataset (Population UNPD) can be run. For this operation open a session as in the previous step but, this time, selecting the Population UNPD dataset and the dimensions (Figure 6).

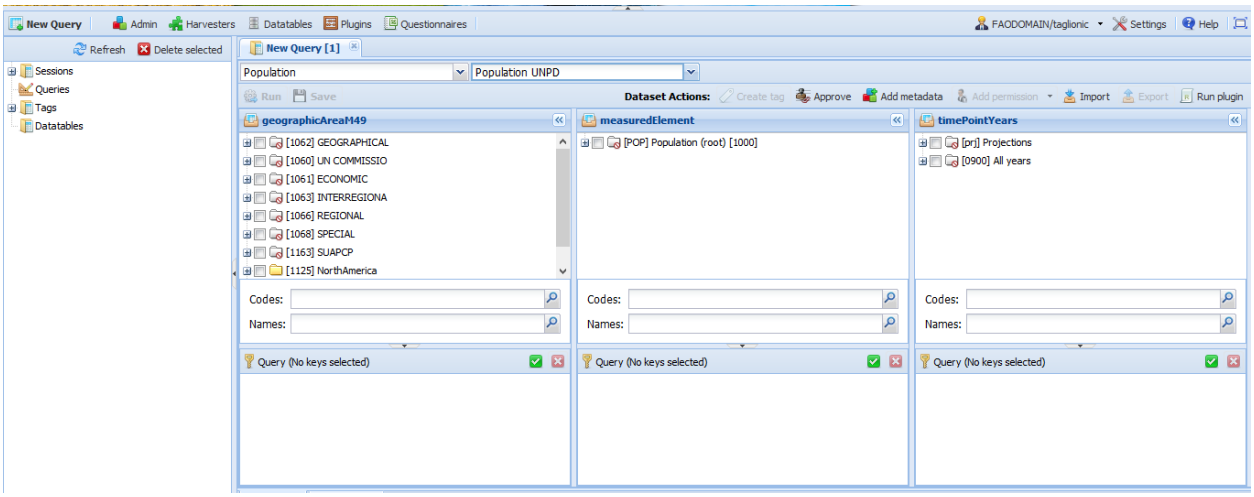

Figure 6

Once the session is open click the 'Run plugin' button in the top right part of the page (Figure 7)

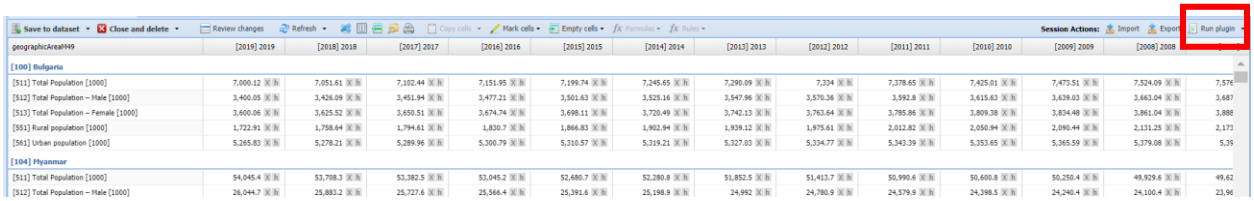

Figure 7

From the list of work select the 'population_unpd2fao' plugin (Figure 8) and then insert the years of the WPP and the WUP used (Figure 9).

When the plugin has finished running an email is sent to the user providing a brief feedback. NOTE, if the plugin fails no email will be sent but just an error will be displayed in SWS. In this case contact the Methodological team.

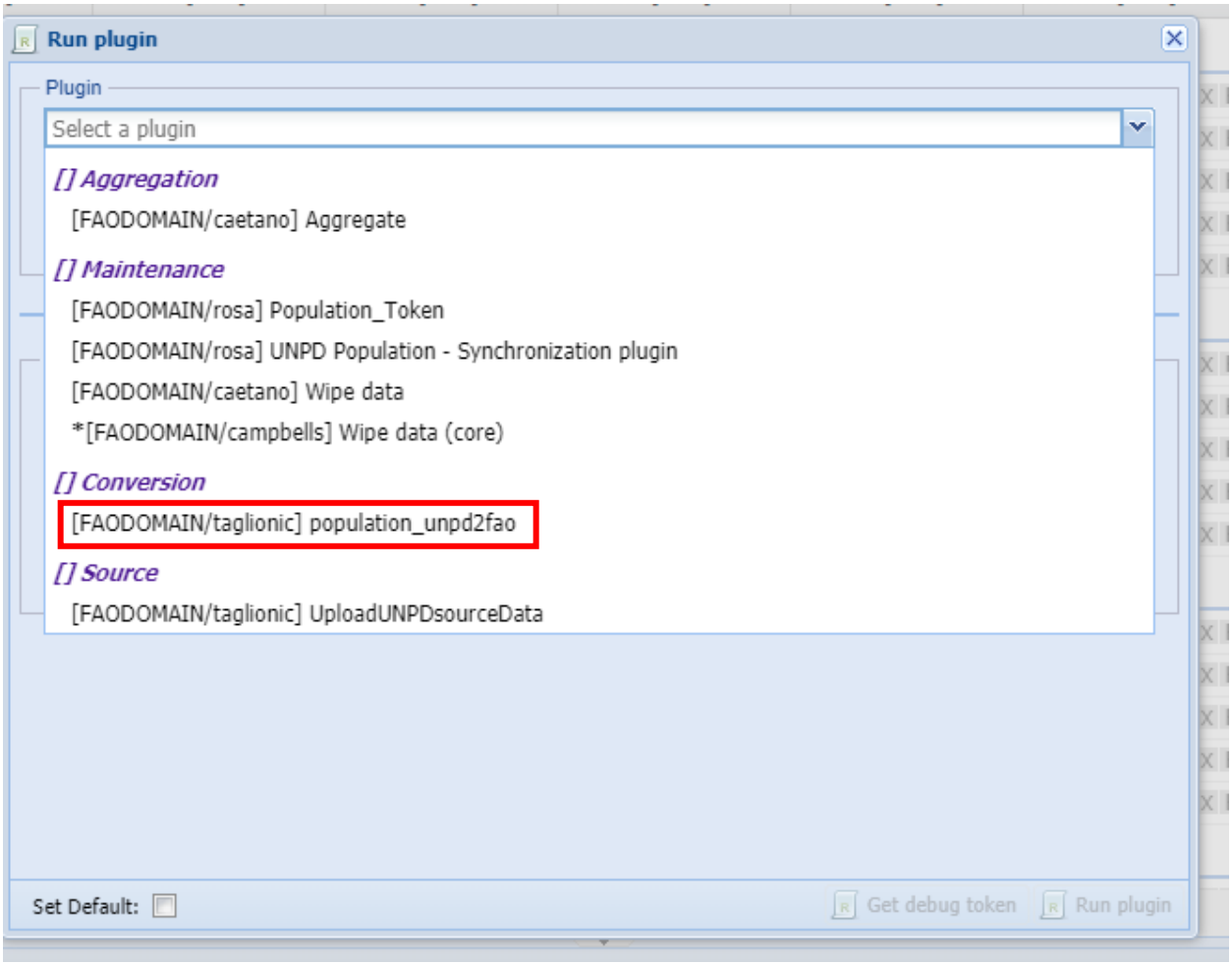
Run plugin
Plugin
Select a plugin
[] Aggregation
[FAODOMAIN/caetano] Aggregate
[] Maintenance
[FAODOMAIN/rosa] Population_Token
[FAODOMAIN/rosa] UNPD Population - Synchronization plugin
[FAODOMAIN/caetano] Wipe data
*[FAODOMAIN/campbells] Wipe data (core)
[] Conversion
[FAODOMAIN/taglionic] population_unpd2fao
[] Source
[FAODOMAIN/taglionic] UploadUNPDsourceData
Set Default:
Get debug token
Run plugin

Figure 8

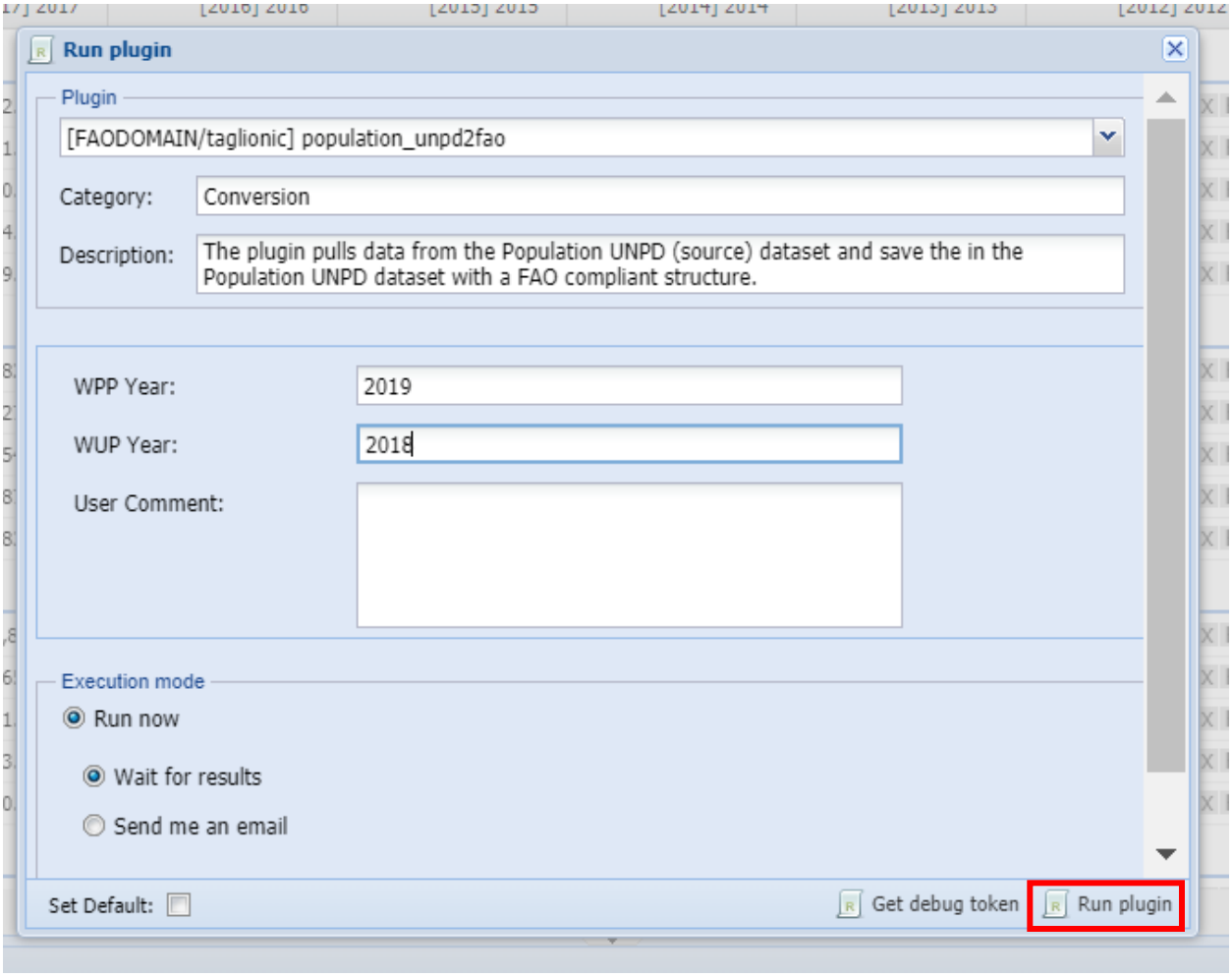

Figure 9

When data have been updated, open again the previously created session of the Population UNPD dataset (session saved appear on the left part of the screen) and after a quick check on changes click on 'Save to dataset' (Figure 10) to permanently commit the changes (the user can equally delete or keep the session when saving).

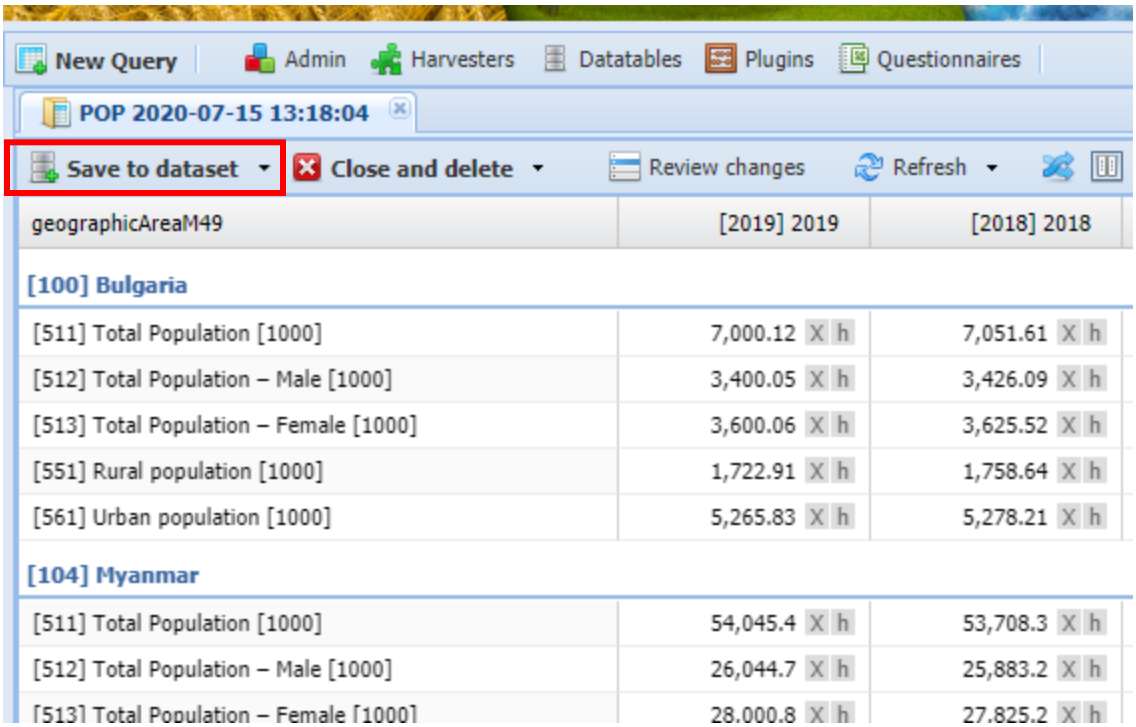

Figure 10

After this step, the dataset is ready for use and the tag can be created and released.

To create the tag, go as for creating a session of the 'Population UNPD' dataset, select all the dimension and click on the 'Create tag' button (Figure 11). Following the instruction give a name to the tag and save.

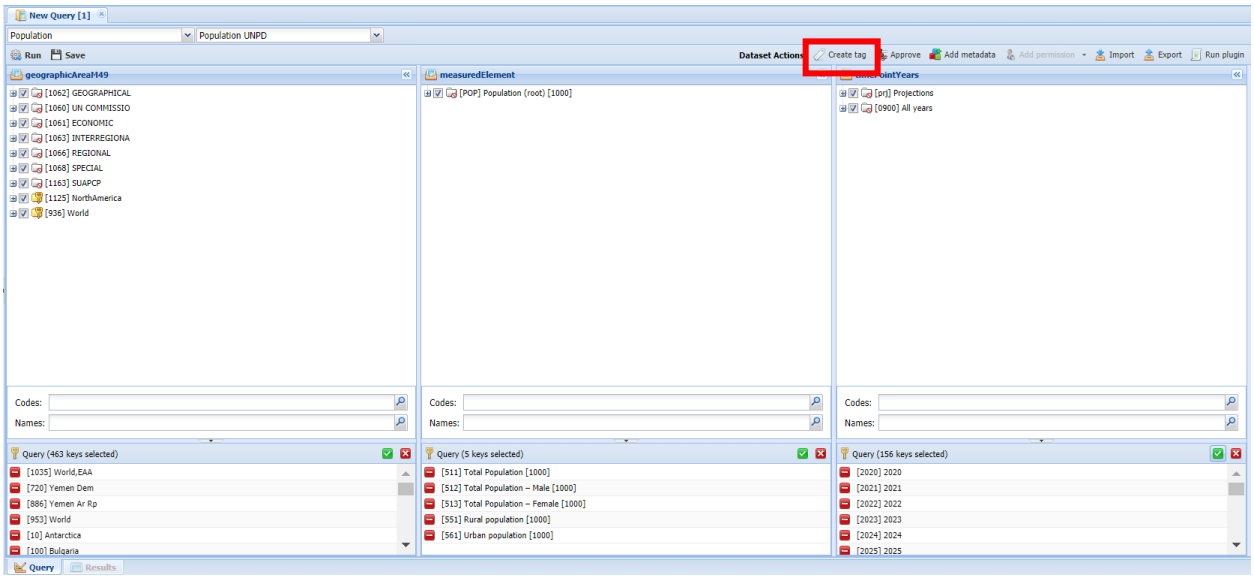

Figure 11

The tag can be found in the menu on the left part of the screen as in Figure 12

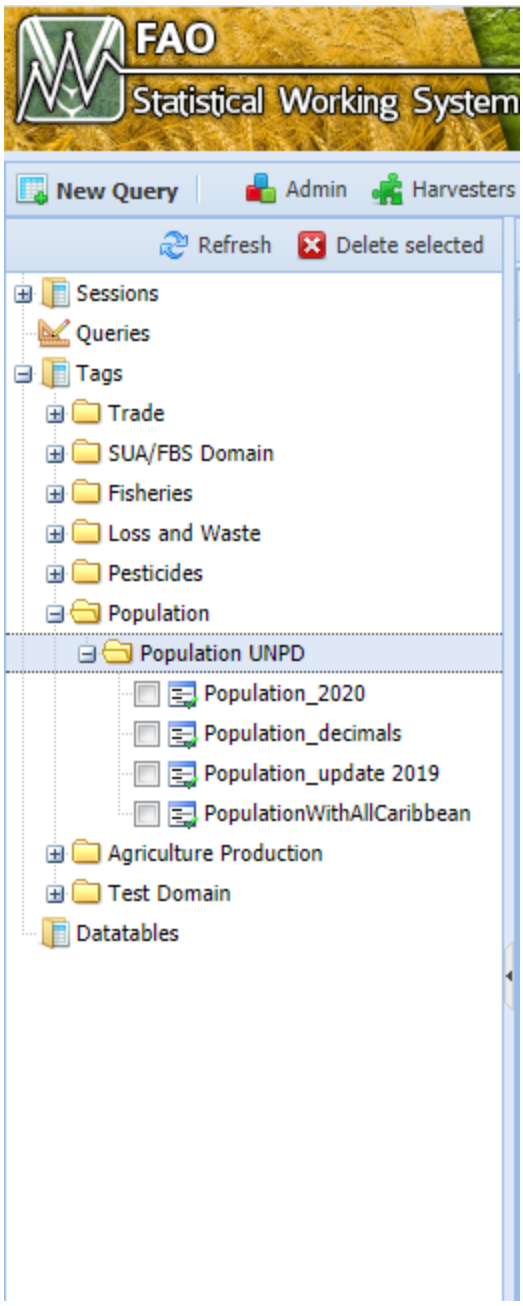

Figure 12

Select the tag just saved, right click on it and select the 'Release' option following the instructions on the screen. The tag is now saved and available to be exported and released on FAOSTAT.

Another operation to perform with the dataset tag is to copy it to the 'Disseminated Datasets' domain (Figures 13-14). This will further ensure all changes committed to this dataset have been reviewed and exactly correspond to what has been published. All the domain using Population data should refer to this dataset.

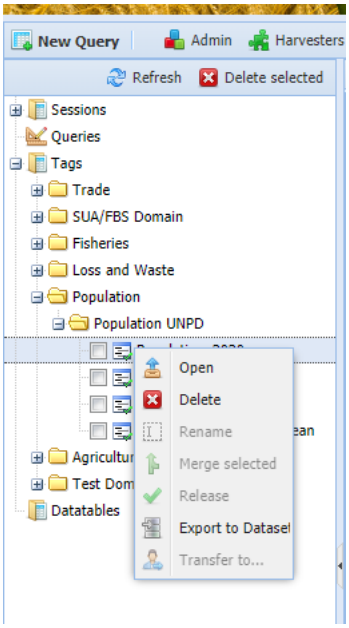

Figure 13

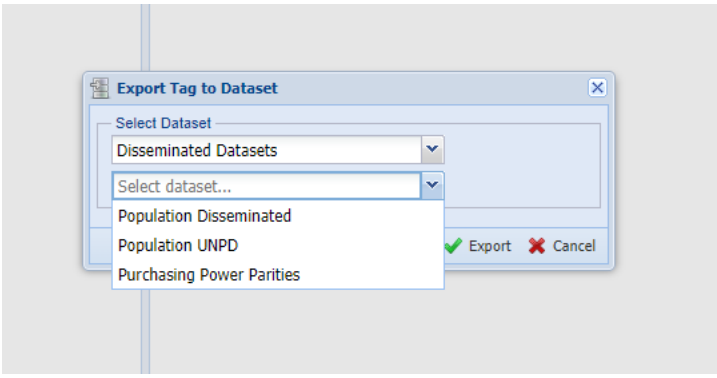

Figure 14

Please note that aggregates are not included in the dataset as aggregates cannot be saved in the dataset, they will be calculated separately as it has been done so far.

After the email confirms data have been copied to the 'Population Disseminated' dataset the whole procedure is complete, **please remember to notify all the SWS users that Population data have changed as they affect several other processes.**

Charlotte Taglioni
Economics and Social Statistics Division (ESS)
Food and Agriculture Organization of the United Nations (FAO)
Viale delle Terme di Caracalla 00153 Rome, Italy
E-mail: charlotte.taglioni@fao.org