

# How to correctly run the Processed Item Imputation Plugin?

**Francesca Rosa**

Food and Agriculture Organization  
of the United Nations

---

## Abstract

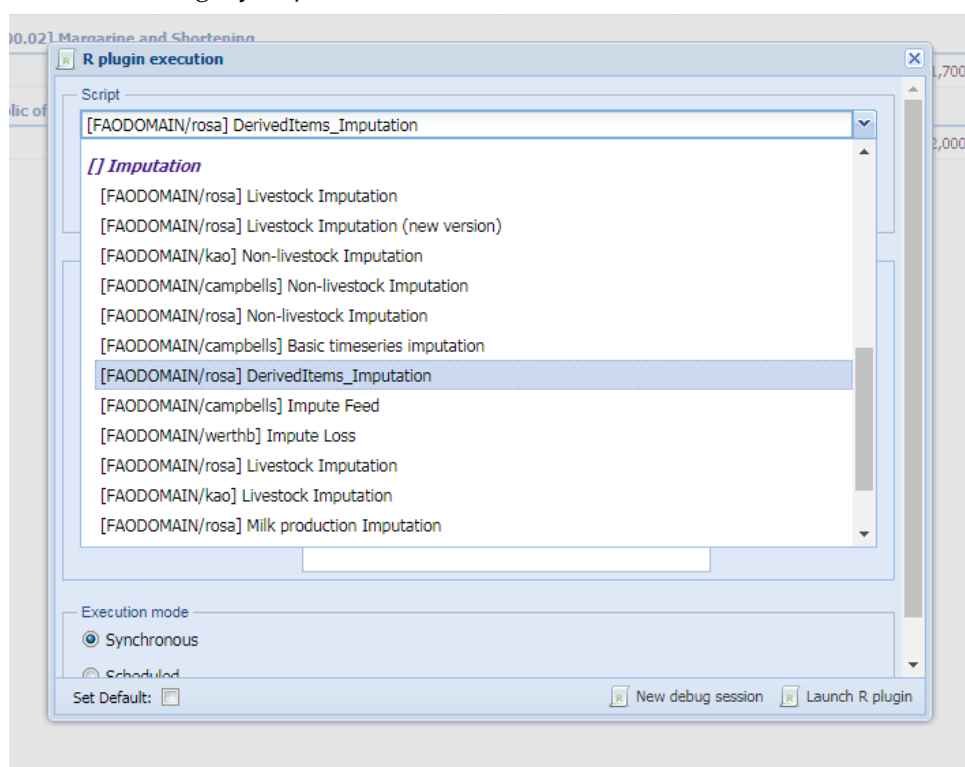
This document has been written to ensure that each user is able to autonomously run the plugin. The first session explains how to set the requested parameters and clarifies the meaning of each requested parameter. The second session makes reference to those data tables (stored in the SWS) containing for each dimension the complete imputation key. It supports the users in case the complete imputation key has to be enlarged or reduced. It basically gives to the user the possibility to change the complete set of data the module has to work on, without modifying the **R** routines. The third session presents some intermediate files produced by the plugin and stored in a shared folder. Those files (both csv and pdf plots) have been created to support the validation process of the output. Finally, in the last session all the involved *dataset* and *data-table* are listed.

*Keywords:* plugin, guidelines, imputation key.

---

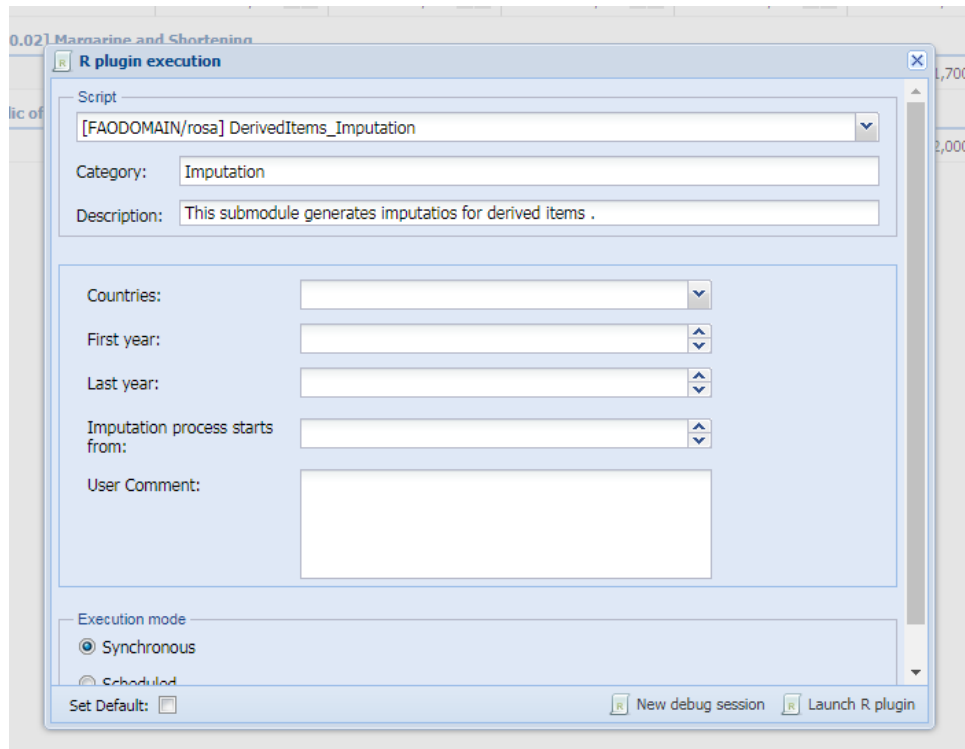
## 1. The plugin

The *Derived Item imputation* sub-module is currently stored in the Statistical working system under the category *Imputation*.



As usual the user is asked to create a new session querying the *Agriculture Production* dataset in the *Agriculture Production* domain.

Choosing the *DerivedItems\_Imputation*, within the list of available plugins, the user is asked to set some basic parameters in order to define the imputation key for the plugin execution.



### 1.1. Countries

The user is asked to declare on which countries the module has to run on. Thanks to this parameter the user can choose if apply the plugin to the countries already contained in the session ("*Country in session*") or to produce imputations for all the country ("*All countries*"). In this latter case the session will be enlarged to include all the FBS countries stored in the data table *FBS\_countries*.

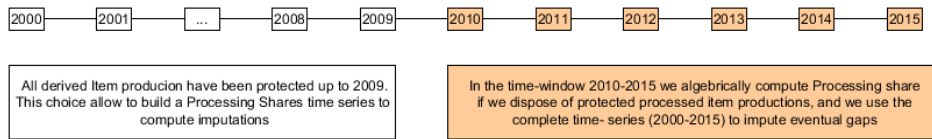
### 1.2. First year, Last year

The user can select the time-window on which the module has to work on. Embedded in the **R** routines there are checks that block the sub-module execution if the user erroneously types unfeasible time windows (e.g. first year greater than last year. )

### 1.3. Imputation process starts from

As already illustrated in depth in the document "*faoswsProduction: A sub-module for the imputation of missing time series data in the Statistical Working System - Processed items*" the module is using a complete time series to impute just a subset of it. In particular, since we do not dispose of broad set of official or semi-official data points, we had to artificially protect some figures. The module splits the time-series in two: the first part is used as training set to produce new imputations for the second part where all the non-protected figures are deleted and overwritten. This parameter controls the break-point. In the example reported in the following figure, we decided to use as training set all the data up to 2009, this means that the *imputation process starts from* 2010. Embedded in the **R** routines there are checks that block the sub-module execution if this parameter do not lie within the time-window specified through

the start year and end year.



## 2. Imputation key

### 2.1. Item

The user has no possibility to choose which set of derived commodity is selected by the module for each country. The processed commodity to be imputed are stored in the SWS data table:

Domain: Agriculture Production	Table: processedItemTable	FAOSTAT
measuredItemC...	measuredItemCPDescription	
23110	Wheat and meslin flour	<input checked="" type="checkbox"/>
23120	Other cereal flours	<input checked="" type="checkbox"/>
23170	Other vegetable flours and meals	<input checked="" type="checkbox"/>
22249.01	Butter and Ghee of Sheep Milk	<input checked="" type="checkbox"/>
22241.01	Butter of Cow Milk	<input checked="" type="checkbox"/>
22254	Cheese from milk of goats, fresh or processed	<input checked="" type="checkbox"/>
22253	Cheese from milk of sheep, fresh or processed	<input checked="" type="checkbox"/>
22110.02	Skim Milk of Cows	<input checked="" type="checkbox"/>
26110	Raw silk (not thrown)	<input checked="" type="checkbox"/>
22251.01	Cheese from Whole Cow Milk	<input checked="" type="checkbox"/>
22251.02	Cheese from Skimmed Cow Milk	<input checked="" type="checkbox"/>
21521	Pig fat, rendered	<input checked="" type="checkbox"/>
22120	Cream, fresh	<input checked="" type="checkbox"/>
22241.02	Ghee from Cow Milk	<input checked="" type="checkbox"/>
22212	Skim milk and whey powder	<input checked="" type="checkbox"/>
22222.01	Whole Milk, Condensed	<input checked="" type="checkbox"/>
22211	Whole milk powder	<input checked="" type="checkbox"/>

The last column, labeled "FAOSTAT", contains a check-box to highlight those items that have to be published on FAOSTAT. The country specific selection of the commodity occurs through the commodity tree which is pulled from the commodity tree dataset and can be browsed in the SWS.

### 2.2. Country

In case the user wants to produce imputations of derived items for all the countries, the module picks this complete list from the data table *FSB\_countries*:

Domain: Agriculture Production	Table: processedItemTable	FAOSTAT
measuredItemC...	measuredItemCPDescription	
23110	Wheat and meslin flour	<input checked="" type="checkbox"/>
23120	Other cereal flours	<input checked="" type="checkbox"/>
23170	Other vegetable flours and meals	<input checked="" type="checkbox"/>
22249.01	Butter and Ghee of Sheep Milk	<input checked="" type="checkbox"/>
22241.01	Butter of Cow Milk	<input checked="" type="checkbox"/>
22254	Cheese from milk of goats, fresh or processed	<input checked="" type="checkbox"/>
22253	Cheese from milk of sheep, fresh or processed	<input checked="" type="checkbox"/>
22110.02	Skim Milk of Cows	<input checked="" type="checkbox"/>
26110	Raw silk (not thrown)	<input checked="" type="checkbox"/>
22251.01	Cheese from Whole Cow Milk	<input checked="" type="checkbox"/>
22251.02	Cheese from Skimmed Cow Milk	<input checked="" type="checkbox"/>
21521	Pig fat, rendered	<input checked="" type="checkbox"/>
22120	Cream, fresh	<input checked="" type="checkbox"/>
22241.02	Ghee from Cow Milk	<input checked="" type="checkbox"/>
22212	Skim milk and whey powder	<input checked="" type="checkbox"/>
22222.01	Whole Milk, Condensed	<input checked="" type="checkbox"/>
22211	Whole milk powder	<input checked="" type="checkbox"/>

At the moment this list contains all the countries for which we dispose of a validated com-

modity tree. The existence of the commodity tree is a necessary condition to ensure that the Derived Item sub-module can be run.

### 2.3. Year

The year selection occurs only through the selection of the plugin parameters, already described in the sections 1.2, 1.3 .

## 3. Validation Output

Since the plugin execution may last several hours (around 5 for all countries), the user receives an e-mail for advising on the successfully execution. Imputed figures are saved in the session and highlighted with a red sign (as usual). To facilitate the validation process some intermediate files are stored in the following shared folder: *hqlprsws1.hq.un.fao.org* .

In particular:

- The *plot* folder contains plots by commodity of all the final time-series obtained by the module. Protected and not protected figures can be recognize by the use of a different symbols (please check the legend on the top of the page). Each chart contains a comparison between the new imputations (*Values*) and the corresponding data-points already stored in the SWS (*oldData*).
- The *output* folder contains a list of csv files (one for each country) containing both the parent SUA and the child new imputed production. In addition it is possible to browse and validate the resulting intermediate coefficients as the *processing shares*, the *availabilities* and the *share down-up* produced by the module.
- The *recovery* folder contains some intermediate csv files that are not useful for validation purposes. These files are stored in the shared folder to be used just in case the module fails and it is necessary to force the the restarting.

## 4. Data sources

This section is not directly useful to physically run the plugin, but it is important being aware about all the datasets involved in the process. In case some outliers or unfeasible figures are produced by the module (e.g. zero values due to negative primary availabilities) the user should know how to manual intervene to quickly take action.

### 4.1. SUA components

Production		Element: 5510
domain	Agriculture Production	
dataset	Agriculture Production	

trade		Element: 5525
domain	Agriculture Production	
dataset	Agriculture Production	

The trade data source depends on the time window considered: in particular from 2013 onwards, data are pulled from the *Total trade (CPC)* dataset, while data upto 2013 are pulled from FAOSTAT 1.

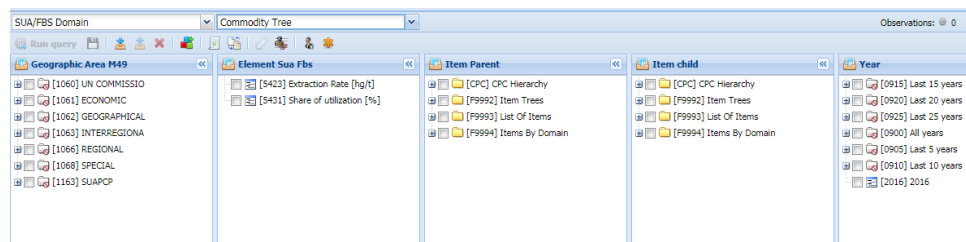
Trade	Elements: 91-61
domain dataset	FAOStat 1 Updated SUA 2013
<i>Trade up to 2013</i>	

Trade	Elements: 5610-5910
domain dataset	Trade Totale trade (CPC)
<i>Trade from 2013 onwards</i>	

Stock Variation	Element: 5071
domain dataset	Stock stocksdata

## 4.2. Commodity tree

The commodity tree is pulled directly from the commodity tree dataset:



## 4.3. The output

The output is save back to the Agriculture Production dataset. In particular only the element-code 5510, corresponding to the *PRODUCTION expressed in tons*, is saved in the session.

### Affiliation:

Francesca Rosa

Economics and Social Statistics Division (ESS)

Economic and Social Development Department (ES)

Food and Agriculture Organization of the United Nations (FAO)

Viale delle Terme di Caracalla 00153 Rome, Italy

E-mail: [Francesca.Rosa@fao.org](mailto:Francesca.Rosa@fao.org)

URL: <https://svn.fao.org/projects/SWS/RModules/faoswsProduction/>