

Food Balance Sheet workflow in the Statistical Working System

Cristina Muschitiello

Food and Agriculture Organization of the United Nations

25 May 2018

Abstract

This vignette provides a description of the workflow and dependencies of operations in the Statistical Working System for the production of Food Balance Sheets.

Contents

Disclaimer	2
Variables of The FBS	2
Definitions	4
The actors of the FBS in the SWS	5
The Overall Workflow of the FBS	5
Flag Management in the FBS framework	9
1 Initial data and the first set of modules run	10
1.1 The <i>Agriculture-aproduction</i> dataset	10
1.1.1 <i>Industrial use Module</i>	10
1.1.2 <i>crop Production Module</i>	10
1.1.3 <i>Livestock Module</i>	10
1.2 <i>Trade Module</i>	10
1.2.1 Module on Crop production	10

List of Tables

1 Valid Flag combinations from Valid Flags table	9
--	---

List of Figures

1 Legend of Objects	5
2 FBS Datasets and their domains in the SWS	6
3 FBS Datatables and their domains in the SWS	7
4 The overall workflow	8

Disclaimer

This Working Paper should not be reported as representing the official view of the FAO. The views expressed in this Working Paper are those of the author and do not necessarily represent those of the FAO or FAO policy. Working Papers describe research in progress by the authors and are published to elicit comments and to further discussion.

This paper is dynamically generated on May 25, 2018 and is subject to changes and updates.

Variables of The FBS

The process of creating FBSs starts by collecting all data for the different variables of the *Food Balance Sheet* equation¹:

$$P_{ijt} + I_{ijt} - X_{ijt} - \Delta St_{ijt} = FP_{ijt} + Fo_{ijt} + Fe_{ijt} + Lo_{ijt} + Se_{ijt} + IU_{ijt} + T_{ijt} + ROU_{ijt} \quad (1)$$

where the i index runs over all countries, the j index over all commodities, and t over years and where, dropping indices for brevity:

- P =Production
- I =Imports
- X =Exports
- S =Stock level
- ΔSt_t = Stock Variation = $St_t - St_{t-1}$
- FP_{ijt} = Food Processing
- Fo =Food availability
- Fe =Feed
- Lo =Losses
- Se =Seed
- IU =industrial use
- T =Tourist consumption
- ROU =Residual Other Use
- $TS = Totalsupply = P_{ijt} + I_{ijt} - \Delta St_{ijt}$

At international level, the primary data source that FAO uses to compile the the Supply Utilization Accounts/Food Balance Sheets are the data as collected through the annual *Agriculture Production Questionnaires*. Unfortunately, measured values are mostly limited to variables on the supply side (production, imports and exports), while, on the demand side, most values are imputed data². The Variables of the FBS are:

- *Production (P)*: Data on production are data at farmgate level. As data on production are very important for countries, these data are very often survey-based data. Nevertheless, not all countries collect data of production for all commodities. Therefore other data collection methods are used, like records of private firms and commodity organization. When no other data are available, Production figures are imputed or estimated. Imputation and estimation procedures depends on the specific commodity. There are different procedures for crops and livestock but all based on an *ensemble approach*. Production data in the FBS framework are collected, imputed or estimated for all the primary commodities and for a set of derived commodities³.

¹For definitions and an extended description of the motivation behind the development of FBS, see FAO, 2001, *Food Balance Sheets: A Handbook*, available at: <http://www.fao.org/docrep/003/X9892E/X9892E00.HTM>. Accessed on 19 January 2017. Moreover see *Standardization & Balancing for Food Balance Sheet Calculation*, in the *Standardization & Balancing* module's documentation on *GitHub*

²For more details on FBS variables please see the latest version of the *Resource Book*

³See *Production module* documentation for more details on procedures and list of commodities

- *Import(I) and Export(X)*: Data on Trade are, mainly, official from international trade databases, like UNSD and EUROSTAT, at HS6 commodity level⁴. Official data are integrated with supplementary data having the main aim of filling in all the information hidden by the unrecorded trade and coming, mainly, from trading partners. HS6 classification is more detailed than the used CPC classification, therefore, these data are aggregated in CPC commodities before being used in the standardization process⁵.
- *Stock Variation (ΔSt_t)*: In the FBS framework, stocks are considered as *changes in stocks* from one time period to the next. Moreover they are considered as a component of supply. Therefore, the $-$ sign indicates that the stock is decreased, which means that the stocks are available as a supply, while the $+$ sign indicates that the stocks have increased and they are, therefore, considered as a utilization for that commodity. Changes in stocks are typically limited to a short number of commodities, mainly grains, pulses and sugar and, because they are very rarely measured by country, figures are very often imputed or estimated⁶. Estimation of *changes in stock* is based on opening stocks figures through an approach that maintain time consistency of available data and official data⁷.
- *Food availability (Fo)*: Food availability is defined as the quantity of any substance that is available for human consumption at the retail level by the country's resident population during a given reference period. Official data of food availability come from questionnaires, industrial output surveys and household consumption or expenditure surveys. When these sources of data are not available, food availability data are imputed or estimated. Not all CPC commodities are Food commodities, as not all commodities are used for human consumption. Food commodities are divided in two main groups: *Food Estimates* and *Food Residual* and estimated differently depending on the pertaining group, respectively as linear or logarithmic function of income elasticity of demand, GDP per capita, and population, or as residual quantity of production and net trade quantities⁸.
- *Feed (Fe)*: Feed demand is increasing because of the increase in income of developing countries. Animal feed may vary among countries due to the difference in livestock and the diversity of commodity used for livestock's rations. Official feed demand data might be available from specific questionnaires. Even when available, these data need to be cross-checked against livestock availability in terms of requirements. When official data, and also other sources of semi-official data, are not available, feed data are estimated as a function of livestock availability and livestock feed demand in terms of energy and protein requirements, in accordance with an inventory of the potential feed supply's products of any country.
- *Seed (Se)*: Official seed data may come from agricultural surveys, while other sources of data might be found in some technical publication. When data are not available these are estimated as a function of a seeding rate and a sown area in the following year.
- *Tourism Consumption (T)*: Tourism consumption is considered here as a separate utilization variable, while in the past it was included in the "other utilization" catchall category. Official data for this variable are rare and may come from tourism offices or collected by tourism boards through surveys. UNWTO is an alternative source of data, but other authorities might also be used. Imputations and estimations of tourist food are made as a function of food figures.
- *Industrial Use (IU)*: This variable refers to utilization of any food items in any non-food industry. Non-food use of food commodities is growing and is highly context and country specific. For this reason there are not, at the moment, suggestions on how to impute and estimate missing figures. As a

⁴Harmonized Commodity Description and Coding Systems (HS) is an international classification of products held by UNSD. Is made of six-digit level codes and used worldwide for trading data classifications. See official *HS6 UNSD webpage* for more details

⁵See *Trade module* documentation for more details

⁶For a complete list of stock commodities and for details about the imputation methodology, please refer to specific documentation for stock

⁷All data are marked as *official*, *semi-official* or *unofficial*, depending on the source they come from, through *flags*. Flag management is one of the core responsibilities of the *Office of the Chief Statistician* Department in FAO. Flags are used from all the estimation procedure for distinguishing between different level of reliability in the data. The most reliable data are used to estimate missing or less reliable data.[this has to be better specified]

⁸For a complete list of food commodities of the two typologies and for details about the imputation methodology, please refer to specific documentation for Food availability

consequence, Industrial data available for Food Balance Sheets are only those coming from Official or unofficial sources. At the moment the data used comes from USDA and from questionnaires.

- *Loss (L)*: FAO has developed the Global Food Loss Index (GFLI) that focuses on the supply-side aspects of improving the efficiency of global food supply chains. The index is based on a set of primary commodities that are key in agricultural production systems, including crops, livestock, and fisheries. In order to track losses without compounding production variability, losses are expressed as a percentage and are aggregated using fixed quantities and prices. The primary data source that FAO uses for compiling GFLI are loss factors as collected by Questionnaires. Other sources are publications and reports from subnational reports, academic institutions, international organizations and so on. The missing data are imputed using a hierarchical model based on commodity groups⁹.
- *Residual and other use (ROU)*: ROU is used to capture categories of products that do not follow in any other category and that might be considered “not important” for the FBS scope. Normally, these residual commodities are different from country to country and for this reason they fall in this variable. ROU are set not to be higher than 5% of total supply and are calculated at-post as absorbing element, in the sense that it absorbs part or all of the imbalance that may exist, at FBS commodity aggregate level, after the standardization process. Any imbalance bigger than 5% of total supply is balanced through a balancing mechanism that will be later specified.
- *Food Processing (FP)*: This variable represents the amount of the availability of a commodity that enters a manufacturing process to be transformed into a derived commodity. Food processing is not officially measured, nor collected via official sources. This variable is entirely calculated during the standardization process by applying extraction rates to the amount of production of the derived commodity. This will be better specified in section 2-Step2 of the present document.

The data of each variable are generally checked and imputed in time series. The set of operations required for creating/checking time series of data for each variable is called *module*. A **module**, in the FBS Framework, is an R-script, written by an R-developer and integrate inside the **Statistical Working System (SWS)**¹⁰ by means of *plugins*. There is at least one module (there might be more) for each variable of the FBS. Each module produces figures that are collected in a dataset inside the SWS for future uses or publication. Output data of a module may become input data of another module, this circumstance creating a precise sequence for the execution of a complete FBS.

Definitions

there are 8 main typologies of objects that will be use in the present document. These correspond to objects that are used in the SWS in managing processes:

1. **Domain**: A domain is an area of work where other objects are grouped, by a common criterion of interest.
2. **Data**: In this framework, every information is grouped under this name, which will be characterized by its origin, whether it is collected or created.
3. **Datasets**: A dataset inside the SWS. Each Dataset can belong only to one domain. Datasets allow for:
 - metadata management,
 - detailed data history
4. **Data Tables**: Data Tables are less protected object that can be easily modified and do not keep history of data.
5. **Modules**: a Module is a process, writte, in this context, in R, which performs different typologies of processes:
 - data manipulation,

⁹ask for links to a proper documentation

¹⁰SWS is an internal Working System providing a platform for statisticians and statistical clers to collect, collate, validate and correct data. Moreover, the platfors supports the possibility of performing imputations of data based on statisticians' knowledge and development.

- statistical analysis,
 - other kind of operations on data.
6. **Plugin**: a Plugin is a module when it is integrated inside the SWS and executable from any user.
 7. **Data Flow**: any information's flow or transfer.

For the purposes of this document the following notation will be used:

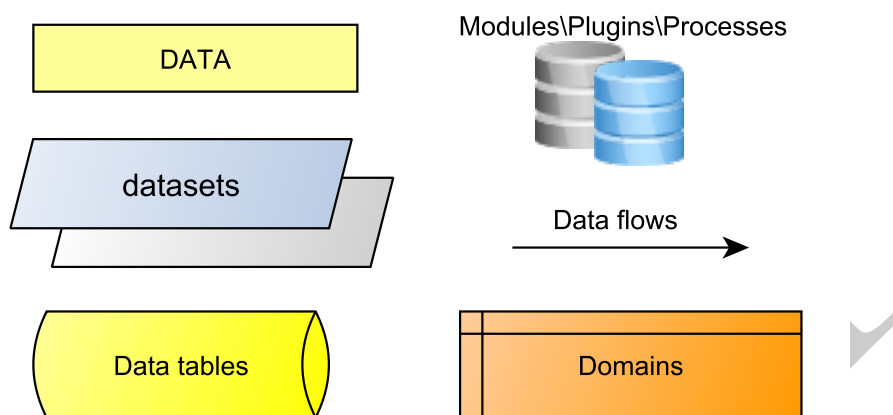


Figure 1: Legend of Objects

The actors of the FBS in the SWS

In the compilation of FBS involves 11 domains:

1. Agriculture production
2. Trade
3. trade-input-data
4. Stock
5. Industrial Use
6. Food Domain
7. Loss and Waste
8. Tourism Domain
9. SUA/FBS Domain
10. United States Department of Agriculture
11. Population
12. faostat_datasets

Inside these domains 18 datasets and 11 data tables are involved, as represented in figures Figure 2 and 3. Details on these objects will be provided along the document.

The Overall Workflow of the FBS

The workflow for compiling Food Balance Sheet is articulated and has dependencies, as some module to run properly might use data that are the output of other modules.

Figure 4 presents a simplistic representation of the overall Workflow. As shown, it consists of 8 propaedeutic steps that will be described in this document.

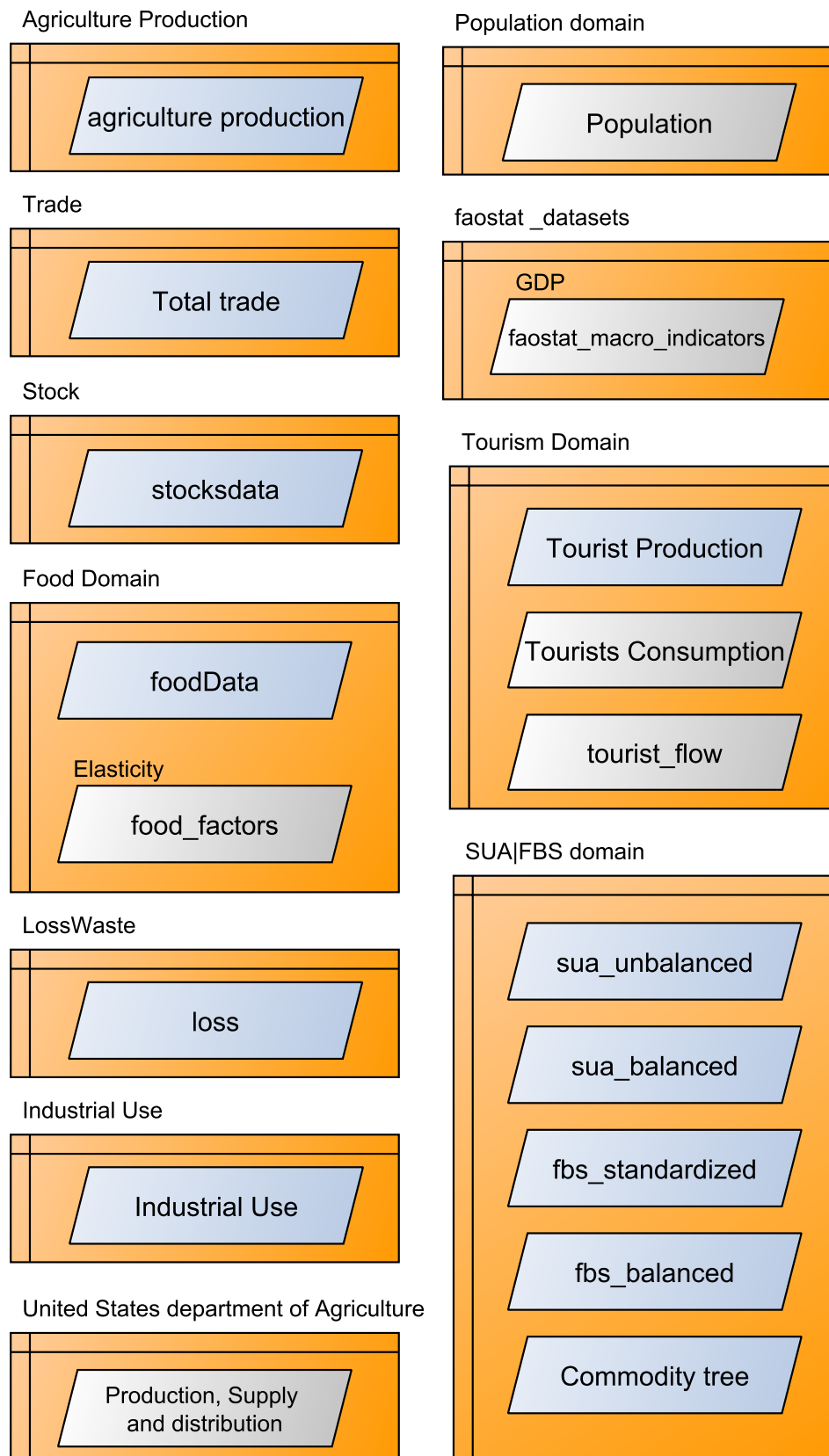


Figure 2: FBS Datasets and their domains in the SWS

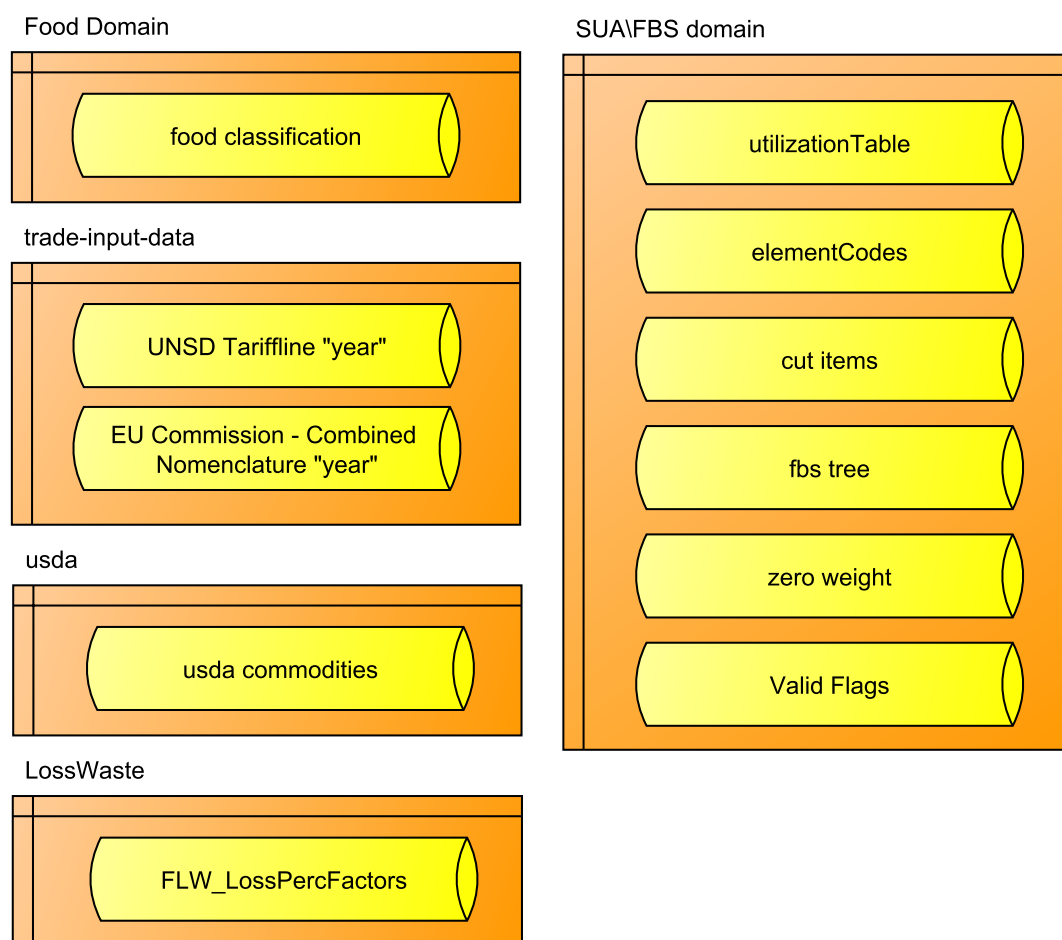


Figure 3: FBS Datatables and their domains in the SWS

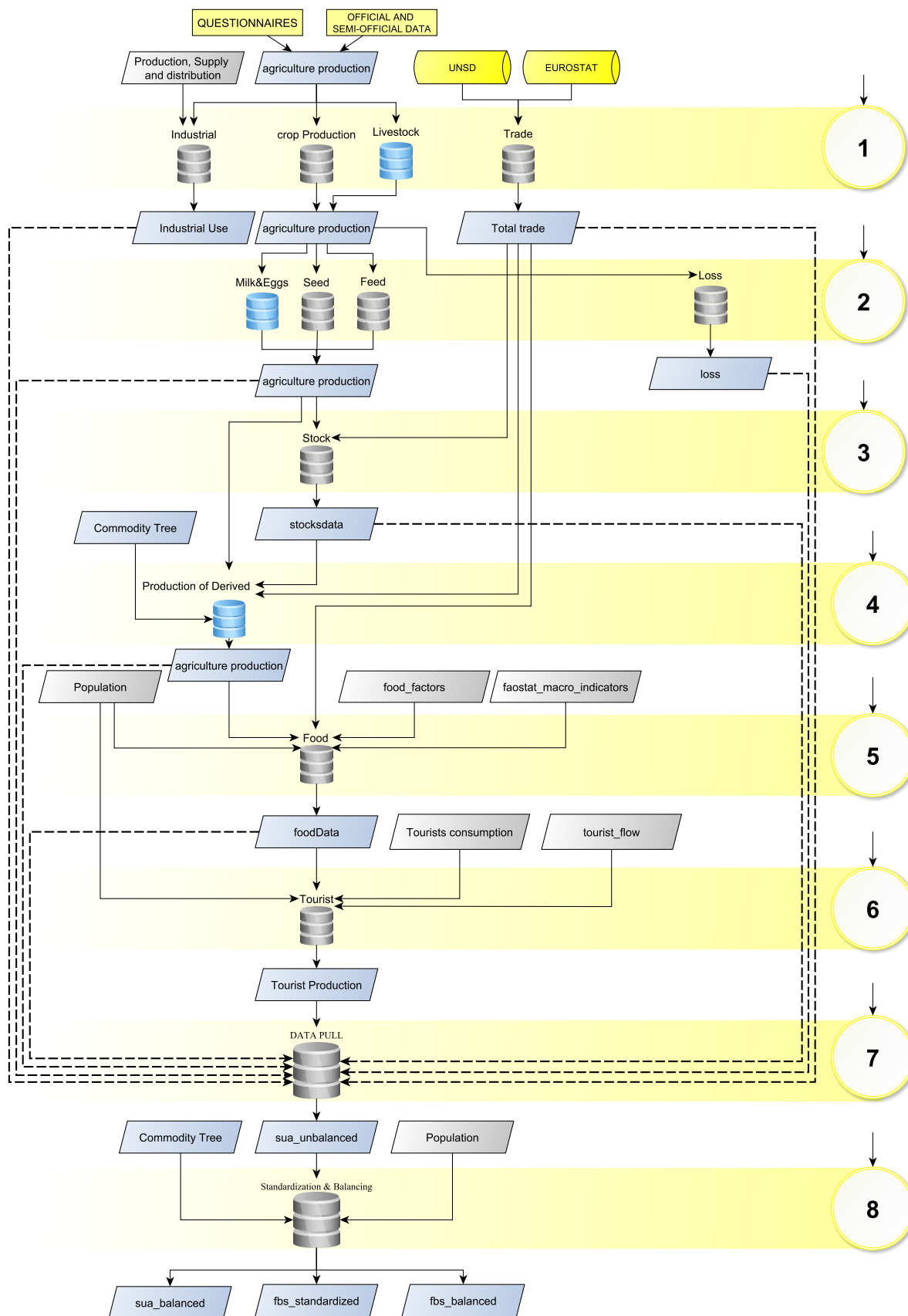


Figure 4: The overall workflow

Flag Management in the FBS framework

Distinction between sources of data in the overall process of producing FBS is performed through Flags. Flags in FAO play a fundamental role and are managed by the *Office of the Chief Statistician* Division in FAO¹¹. In the context of FBS flags are used accordingly to a table (stored in the SWS as Data table in the SUA/FBS domain) which set the combination of flags that are associated to valid, invalid, official, unofficial estimated data, etc. Table 1 reports the Valid Flag combination in the SWS, taken from this Data Table.

Table 1: Valid Flag combinations from Valid Flags table

flagObservationStatus	flagMethod	Valid	Protected
E	-	TRUE	FALSE
E	c	TRUE	TRUE
E	f	TRUE	TRUE
E	h	TRUE	TRUE
E	i	TRUE	FALSE
E	s	TRUE	FALSE
I	-	TRUE	FALSE
I	b	TRUE	FALSE
I	c	TRUE	TRUE
I	e	TRUE	FALSE
I	i	TRUE	FALSE
I	s	TRUE	FALSE
M	-	TRUE	FALSE
M	c	TRUE	FALSE
M	i	TRUE	FALSE
M	q	TRUE	FALSE
M	s	TRUE	FALSE
M	u	TRUE	FALSE
T	-	TRUE	TRUE
T	c	TRUE	TRUE
T	h	TRUE	TRUE
T	i	TRUE	FALSE
T	p	TRUE	TRUE
T	s	TRUE	FALSE
	-	TRUE	TRUE
	c	TRUE	TRUE
	h	TRUE	TRUE
	i	TRUE	FALSE
	p	TRUE	TRUE
	q	TRUE	TRUE
	s	TRUE	FALSE

A specific agreement exists in this context for which all data coming from old methodology, which are often used for consistency and coherence with the past, are assigned a *method flag* equal to “-”.

¹¹link some documentation on Flags... which one?

1 Initial data and the first set of modules run

1.1 The *Agriculture-aproduction* dataset

The data collected through the *Agriculture Production Questionnaires* and from other official or semi-official sources are collected inside the *Agriculture:aproduction* (i.e. *agriculture domain, aproduction* dataset) dataset. Here all data are collected from countries on:

- Area harvested
- Yield
- Area sawn
- Carcass weight
- Livestock
- Milking animals
- Milking products
- Seed qt
- Crop Production qt
- stocks qt
- Industrial Use qt
- Feed qt
- Food availability qt
- Losses qt

It represents the main input dataset containing input-data of almost all domains (Trade is excluded as it takes data from different sources). It has a special role in the all FBS framework and its content and use is very delicate and not straightforward. In particular this dataset contains different typologies of data, some used as they are and other used for producing other data:

1. Data on area harvested, yield, area sawn are used for estimating missing production data and seed data;
2. Data on Crop Production, Industrial Use, Feed, Seed, Food and Losses originally collected in this dataset are *Protected*, which means that they will be used for:
 - estimating missing production data,
 - estimating missing seed data,
 - estimating missing feed data,
 - directly used for the FBS calculation
3. Data on Carcass weight, livestock, milking animals and milking products are used by other modules for producing data on meat, milk and eggs.

Indeed, three modules are run which have this dataset, as it is at the beginning of the overall process, as, unique or not, input:

1. Industrial use Module
2. crop Production Module
3. Livestock Module

1.1.1 *Industrial use Module*

1.1.2 *crop Production Module*

1.1.3 *Livestock Module*

1.2 *Trade Module*

1.2.1 Module on Crop production