

Food Balance Sheet workflow in the Statistical Working System

Cristina Muschitiello

Food and Agriculture Organization of the United Nations

24 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
Objects in the SWS (DEFINITIONS)	4
1 The Overall Workflow of the FBS	5
Commodity Tree	6
2 Data Pull	8
the Initial Sua Unbalanced	8

List of Tables

1 Unbalanced Sua table - China/Wheat/2014 example	9
---	---

List of Figures

1 Legend of Objects	5
2 Standardization and Balancing Overall Workflow	6
3 Commodity Tree for Wheat in China mainland 2014	7

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 24, 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.

Objects in the SWS (DEFINITIONS)

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

⁹ask for links to a proper documentation

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

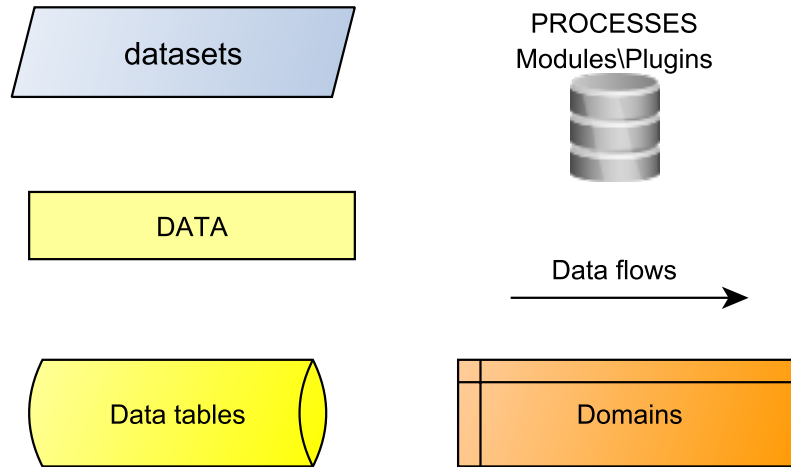


Figure 1: Legend of Objects

1 The Overall Workflow of the FBS

The *Standardiation & Balancing* process is presented in Figure 2. It involves 5 main steps and requires a some auxiliary information table.

The 5 steps are:

1. Data Pull,
2. Sua Filling,
3. Standardization,
4. Balancing,
5. FBS aggregation.

The additional information's tables are:

- *Utilization Table*,
- *Zero-Weight* table,
- *cut* table,
- *Fbs Tree*.

All the steps and the table will be described across the present document, after having introduced the the basic equation of FBS and the *Commodity tree* which represents the structure of the set of commodity involved in the Food Balance Sheets.

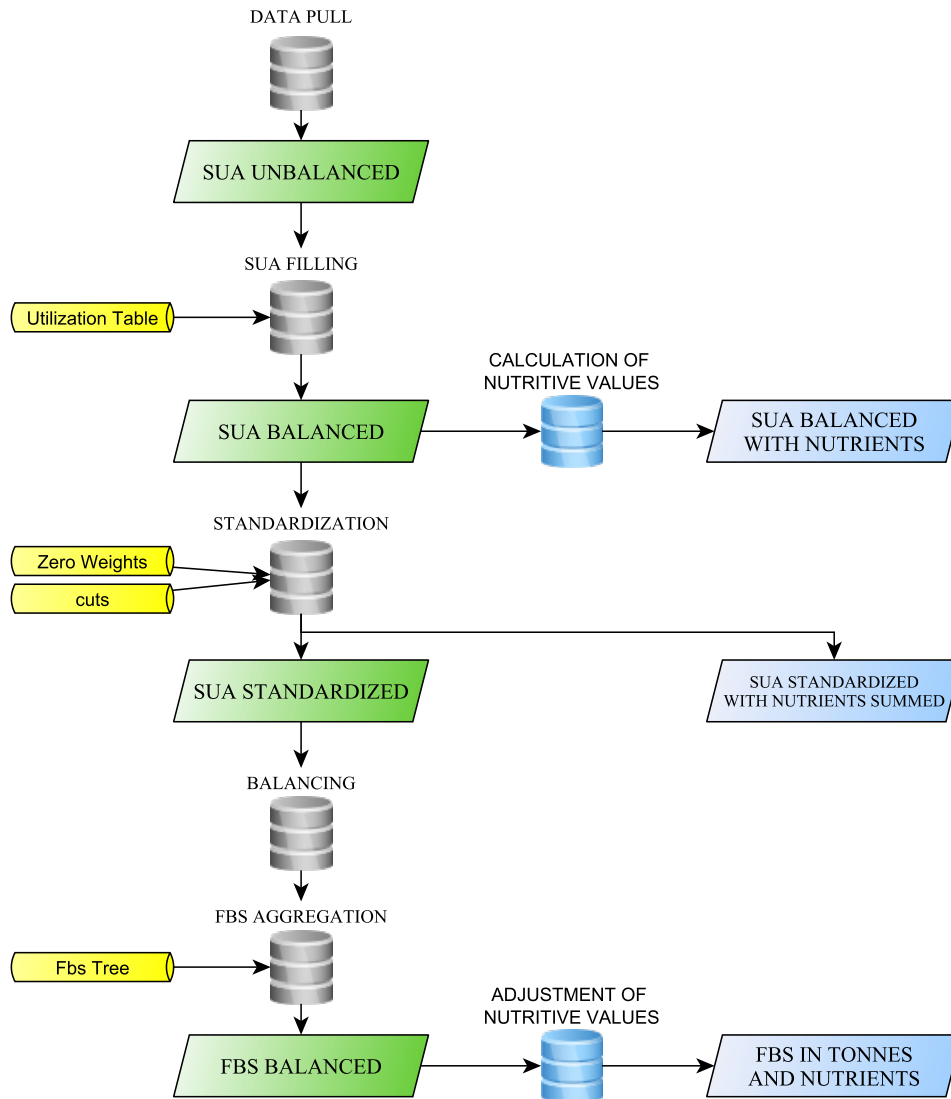


Figure 2: Standardization and Balancing Overall Workflow

Commodity Tree

The process of combining commodity balances for creating Food Balance Sheets is based on a structured and clear set of relationships between commodity given by the *Commodity tree*. The majority of the commodities are produced from one (or more) commodity (/ies), called *parent* commodity(/ies), and/or are themselves parent of one (or more) *child* (*children*) commodity (/ies). These structure creates an intense and articulated network of relationships at different levels: primary commodities, like crops, are *parent* commodities and, also, *zero-level* commodities from which *children* commodities of *level-1* are produced, which are in turn, used to produce other commodities of a gradually “lower” level. In commodity trees, the bigger the level number, the lower the processing level. There are as many commodity trees as the number of process chains in a country. Fundamental characteristics of commodity trees are¹¹:

¹¹For a more detailed description of commodity trees, please see the specific documentation. The reference document, at the moment is the *technical conversion factor* document available in the documentation folder on *GitHub*

- Each commodity tree is represented as a flowchart of the kind presented in Figure 3 where:
 - nodes** represent commodities,
 - edges** represent production processes ,
 - joints** indicate where a single production process creates more than one commodity. These commodities are, then, called *by-products* or *co-products*.

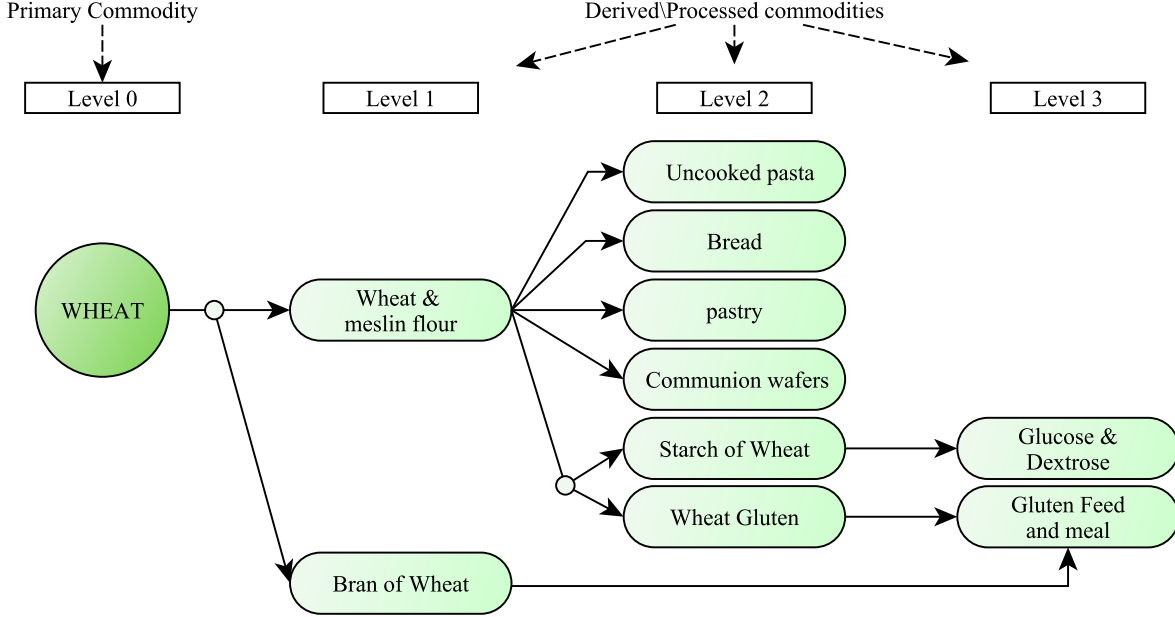


Figure 3: Commodity Tree for Wheat in China mainland 2014

- Not all the countries have the same production processes, because countries have different technologies and primary products availabilities. Therefore, the commodity trees are not the same across countries.
- if a production process is active in a country, this is expressed through the existence of a conversion factor called **extraction Rate**. An Extraction rate (eR) represents how much amount of the child commodity is produced from 1 unit of parent commodity. It is expressed as a ratio of the processed product obtained from the processing of the parent/originating product.
- Some child commodities can be produced starting from more than one parent commodity. A second conversion factor exists representing the amount of a child commodity that is produced from each parent commodity. This conversion factor is called **Share**. Shares represent the amount of the child commodity that is produced from the specified parent and are expressed as a ratio. Shares are generically defined as:

$$s_{cp} = \frac{availability_{p(c)}}{\sum_{p=1}^A availability_{p(c)}} \quad (2)$$

where $availability_{p(c)}$ is the availability of each parent p of child c expressed in terms of c (in *child equivalent*).

Commodity trees are presented in tables like Table 10, which represents the same example of Figure 3. In the table each production process is represented in a separate row.

There are some concepts linked to the *Commodity tree* framework:

- **Proxy-Primary** commodities. These are a set of commodities that are children of other commodities but, because they are important in representing the food availability of a country, are not aggregated to their primary commodities, but are kept separated. These commodities are *cut* from the tree of the primary commodity/ies and, if they can be processed in other products, have their own commodity tree. The name *proxy-primary* is assigned because they are considered as primary-commodities in the *Standardization & Balancing* process.
- **No-Tree** commodities. These are *zero-level* commodities that are primary commodities and are never processed in other products. As they are not involved in any production process, there is no tree associated to them. Notice that, even a commodity that is included in the commodity tree of one Country, might be a No-tree commodity for another country or another year, if no production processes have been activated for that specific Country or year.

2 Data Pull

The process of creating FBSs starts by considering the initial commodity balance for each CPC commodity, either primary and derived, with the different variables of the equation (as listed and briefly described in the previous section) filled with figures as available from official or other sources and from imputation and estimation methods, when applied. In other words, the process starts by pulling figures inside a so-called *Sua Unbalanced*. In this initial account, food processing and ROU figures are not available (because they, by default, will be measured during the process), whereas the figures for all other variables have been already collected, imputed and estimated through a specific *module* (Figure 1).

A **module**, in the FBS Framework, is an R-script, written by an R-developer, for the execution of a set of operations (either data import, manipulation, imputation or estimation) required for compiling the time series of one variable. 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 **Statistical Working System (SWS)**¹². 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. In the present document, we are analyzing the workflow of the Standardization process as starts after all the modules have run and have produced reliable data or each variable. The detailed description of the workflow for the execution of all the modules of the FBS will be given in a separate document.¹³

the Initial Sua Unbalanced

After pulling all data, the process of compiling Food Balance Sheets is a non-complete supply-utilization account. The non-completeness of the SUAs is due to different reasons: first, as already said, some variables are not collected, nor estimated before the process begins, second there is the model for industrial use that does not impute or estimate data, but just collects data from different sources, this opening the strong possibility not to have values where they are supposed to exist and, also, not guaranteeing consistency of data over time. Thirdly, modules might, sometimes, fail in the imputation, because of the strong complexity and structural diversity of the input data.

¹²SWS is an internal Working System providing a platform for statisticians and statistical clerks to collect, collate, validate and correct data. Moreover, the platform supports the possibility of performing imputations of data based on statisticians' knowledge and development.

¹³[report the document when it will be available]

Table 1: Unbalanced Sua table - China/Wheat/2014 example

itemName	P	I	X	DSt	Fo	FP	Fe	Se	T	IU	L	ROU
<i>Wheat</i>	<i>126,208,400</i>	<i>2,971,249</i>	<i>957</i>	<i>1,120,565</i>		-	<i>29,181,617</i>	<i>4,277,567</i>		<i>2,985,279</i>	<i>2,713,000</i>	-
Wheat and meslin flo	70,500,000	33,055	188,674		67,300,000	-			-17,345			-
Mixes and doughs for		6,497	38,072		0	-	0		0			-
Other Fructose and S	126,277	3,659	162,324		0	-			0			-
Starch of Wheat	239,816	11,035	40,311			-	172,196			7,919		-
Wheat Gluten	25,580	877	117,373			-	0					-
Communion wafers	13,263	8,796	5,822		16,241	-			-4			-
Uncooked pasta	1,415,692	12,520	22,550		1,405,661	-			-362			-
Food Preparations of		69,686	21,977		47,709	-			-12			-
Bran of Wheat	21,414,279	156,359	2,200		16,500,000	-	4,827,244		-4,252			-
Gluten Feed and Meal	793,740	160,231	529,333			-						-
bread	15,485	2,897	4,210		14,175	-			-3			-
pastry	193,950	89,593	117,630		165,914	-			-43			-

^a P=Production, I=Import, X=Export, DSt=Delta Stock, Fo=Food Availability, FP=FoodProcessing, Fe=Feed, Se=Seed, T=Tourism Consumption, IU=IndustrialUse, L=Loss, ROU=Residual and other uses

DRAFT