**ORANGE BOOK – *PRELIMINARY TEXT* ON CLASSIFICATIONS**

Valentina Ramaschiello

11 August 2015

**DRAFT**

The **FAOSTAT commodity list (FCL)**[[1]](#footnote-1) is the classification of commodities used in FAOSTAT since the 1960’s. Originally it was based on the UN Standard International Trade Classification (SITC)[[2]](#footnote-2). It includes 683 commodities, grouped in 20 chapters (or groups) and covers crops, livestock and their derived products. It excludes agricultural inputs (such as fertilizers, pesticides and machinery), fishery and forest products, for which different classifications and lists are used in FAOSTAT.

The purpose of the FCL is to provide a framework for collecting and analysing data on production and trade and to compile the Supply Utilization Accounts (SUA). Commodities in Food Balance Sheets (FBS) are defined in terms of primary equivalents and follow an *ad hoc* codification.

ESS collects **production data** through a production questionnaire (PQ) that is sent on annual basis to National Statistical Offices and Ministries of Agriculture around the world (supplementary sources are also used for selected commodities and countries). Product lists included in PQs are country-specific, which means that the number and type of commodities can vary from country to country. The generic template of the PQ includes 209 primary commodities (167 crops and 42 livestock) and only 47 processed products(34 vegetable oils and cakes, 4 dried fruit, 6 alcoholic beverages, 3 sugar products). The classification used is FCL (links to the **UN Central Product Classification - CPC -** have been added for the last two data collection rounds as further explained in the next paragraphs).

**Trade data** in ESS are not collected through a questionnaire: countries send to ESS their full trade files in **Harmonized System (HS)** format; data on food and agriculture are then extracted from trade files. Trade data are also received from the UN Statistics Division (Comtrade). Trade data are eventually converted from HS to FCL format. Once trade data are converted to FCL they are combined with production data to compile SUA/FBS. Trade data are also published on FAOSTAT in FCL format.

The FAO Statistics Division (ESS) is revising the classification system in use in FAOSTAT and replacing the FAOSTAT Commodity List (FCL) with CPC “*expanded for agricultural statistics*” for what concerns agricultural commodities. The change in the classification in FAOSTAT is a challenging process and it requires significant resources and a major collaboration effort amongst Divisions in FAO and with other International Organizations. However, efficiency gains are expected in the longer run.

CPC is a comprehensive classification of products*[[3]](#footnote-3)*, in a system of categories that are both exhaustive and mutually exclusive and based on a set of internationally agreed concepts, definitions, principles and classification rules. CPC has a five-level hierarchical structure where each digit provides information on product grouping (Figure 1).

Figure 1: Codes for “Sweet potatoes” in CPC Ver.2 (0122 in FCL)



The latest CPC version (2.1) will be complemented with an official annex developed by FAO to meet the needs of agricultural statistics; such structure is called “*CPC expanded for agricultural statistics*”. CPC expanded provides additional detail on agricultural commodities (primary products) and is obtained adding one level (two digits) at the lower of the standard CPC[[4]](#footnote-4) (Figure 2).

Figure 2: CPC expanded code for “Mixed grain” (0103 in FCL)



CPC classifies products based on the *physical properties and the intrinsic nature* of the products as well as on the principle of *industrial origin* (harmonized with ISIC - although in some cases products can be the output of severalISIC industries). **HS** subheadings are used as building blocks for the goods part of the CPC: high harmonization with the HS is therefore ensured.

CPC is a general-scope classification i.e. covers products of all economic activities (not sector specific) but it can be customized for sectoral applications. It is also a general-purpose classification, so that potential applications range from production, to trade, prices and consumption.

Benefits from using CPC:

* an international classification, constantly updated and reviewed by the Expert Group on International Classification (chaired by UNSD, participants are countries and international organizations)
* used by other organizations and statistical domains: allows data comparability across statistical domains
* used by countries: reduce reporting burden; in addition CPC expanded is designed not only for FAO but also for countries engaged in the collection and dissemination of data on agriculture and food products: it provides a flexible tool that allows increase granularity at the lower level, including local species and varieties, while maintaining comparability across countries at the higher level
* alignment with ISIC and HS: when these classifications are updated, the CPC is also updated; as it is highly aligned with HS, data conversion for SUA/FBS is improved compared to FCL (65% of the HS07-CPCver.2 correspondences are one-to-one or many-to-one vs. 35% in the case of HS-FCL) (Figures 3 and 4)

Figure 3: HS – FCL correspondences



Figure 4: HS – CPC correspondence



CPC implementation in the production domain

A precondition to the implementation of CPC was to increase detail on agriculture, forest and fishery products in CPC. To this end, FAO contributed significantly to CPC ver.2 and ver.2.1. together with UNSD and the Expert Group on International Classifications. In addition, a CPC expanded for agricultural statistics was developed and added as an official annex to CPC version 2.1. When detail in CPC 2.1 expanded is not sufficient, the classification is expanded further in FAOSTAT according to FAO needs and data available. In this way high harmonization between FCL and CPC is ensured.

CPC is planned to be used for future data collection and to be applied to old time-series, in order to allow data comparability over time and avoid breaks in the series.

As mentioned, countries provide production data to FAO on annual basis, by means of a questionnaire sent by ESS to NSOs and MoAs. In the PQ agricultural products are identified and coded according to the FCL: in view of the change in the classification correlations to CPC codes have been introduced for the last two rounds of data collection.

Although the basic condition for data back cast is to have double coded data for at least one year, it seemed difficult for FAO to increase its data request to countries: additional burden on national offices might have lowered the response rate and hampered the data collection process. Therefore ESS identified alternative solutions to allow progress in the change of the classification and data back cast, while reducing the cost of this operation. The solution adopted depended on the type of link encountered and allowed **full alignment between FCL and CPC**:

* **One-to-one** cases are resolved quite easily as old data are transferred to the new classification assigning codes and definitions according to the new classification while data remain the same (“key method” [[5]](#footnote-5)).
* Also for **many-to-one** cases data conversion is straightforward as data in FCL are aggregated into the target classification (CPC). Such an aggregation entails a loss of information, as CPC is less detailed than FCL. Not to lose information in FAOSTAT, many-to-one cases have been turned into one-to-one correlations: first the target classification is expanded further according to the detail available in FCL and then the “key method” is applied. When detail in CPC 2.1 expanded is not sufficient, the classification is expanded further for FAOSTAT purpose.

More difficulties are faced for one-to-many and many-to-many types of links. In these cases data are converted based on statisticians’ best judgment according to the *dominant* correspondence. Coefficients of conversion have not been calculated, given the lack of information in both formats for at least one year, therefore the risk to threaten data quality in the conversion. Conversion keys used are 1 and 0 exclusively:

* **One-to-many** relations between FCL and CPC are managed identifying the dominant correlation based on statistician’s best judgement and assigning the conversion key “1” accordingly.
* In **many-to-many** cases, which represent a minority in the FCL-CPC correlations, the target classification is modified and aligned to the source one.

Details and examples are provided in Appendix 1.

CPC implementation for SUA and FBS

The compilation of SUA and FBS is based on commodity trees. What is called a “commodity tree” in FAO should not be confused with a classification tree or “hierarchy”.

A **commodity tree** (CT) is a “*symbolic representation of the flow from a primary commodity to various processed products derived from it, together with the conversion factors from one commodity to another*”[[6]](#footnote-6).

A **statistical classification** is “*a set of categories which may be assigned to one or more variables”* where “the *categories are defined in terms of one or more characteristics of a particular population of units of observation. A statistical classification may have a flat, linear structure or may be hierarchically structured, such that all categories at lower levels are sub-categories of a category at the next level up.*”[[7]](#footnote-7)

The FAOSTAT commodity list is a flat classification (or “a list”) where commodities are listed following an ascendant order (in most cases). The FCL itself does not set the relations amongst commodities as all categories are on the same level: to distinguish primary from processed products the printed version of the FCL uses capital letters, which is not a classification feature (Figure 5). It is the commodity tree that sets the links amongst commodities listed in FCL through the application of extraction rates. Extraction rates “*indicate, in percent terms, the amount of the processed product concerned obtained from the processing of the parent/originating product, in most cases a primary products*”[[8]](#footnote-8) (Figure 6).

Figure 5: Classification of millet and its derived products in FCL

**Group 1: Cereals**

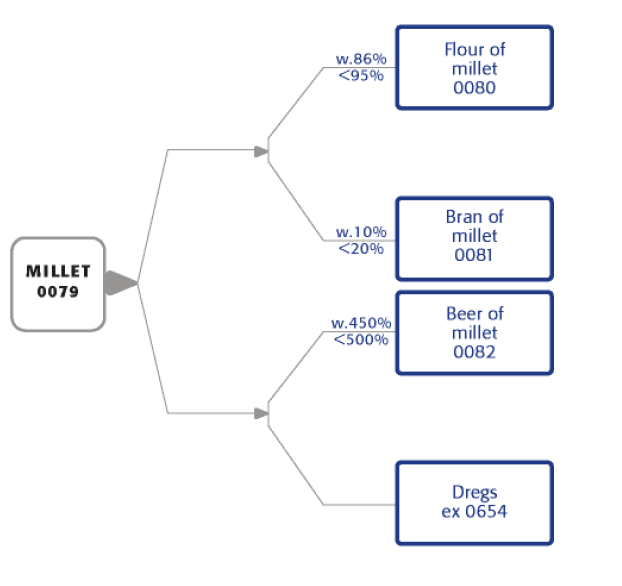
0070 MILLET

0080 Flour of millet

0081 Bran of millet

0082 Beer of millet

Figure 6: Commodity tree of millet (a simple example as it only includes 1st level processed products)



As far as single commodities are identified in the reference product classification (one-to-one and one-to-many correspondences), commodity trees can be developed. Commodity trees are “independent” from the statistical classification used, or better: their structure does not depend on the reference classification hierarchy. Indeed, relations set in the trees should not be confused with the classification hierarchy. In a hierarchical classification items at the lower level can be grouped/aggregated into the one at the higher level. For example: millet, wheat, barley and maize can be grouped into the group “cereals” or seeds and grains of milled can be grouped into the class “millet”. This is not true for commodity trees where flour, bran and beer cannot be grouped into millet unless quantities are first expressed in terms of primary equivalents, by applying extraction rates. Therefore in CT the key is the *relation* set amongst commodities while commodities are listed in the reference product classification (Fig. 7 and 8).

Commodities in FBS are defined in terms of primary equivalents and follow an *ad hoc* codification while links can be made to both CPC and FCL.

Figure 7: Classification of millet and its derived products in CPC; CPC hierarchy reflects the economic activity of origin

**Section 0: Agriculture, forestry and fishery products**

Division 01: Products of agriculture, horticulture and market gardening

Group 011: Cerals

Class **0118: MILLET**

**Section 2: Food products, beverages and tobacco [...]**

Division 23: Grain mill products, starches and starch products; other food products

Group 231: Grain mill products

Class 2312: Other cereals flour

Subclass 23120: Other cereal flours

FAO Expansion **23120.05 Flour of millet**

Division 24: Beverages

Group 243: Malt liquors and malt

Class 24310: Beer made from malt

FAO Expansion **24310.03 Beer of millet**

**Section 3: Other transportable goods**

Division 39: Wastes or scraps

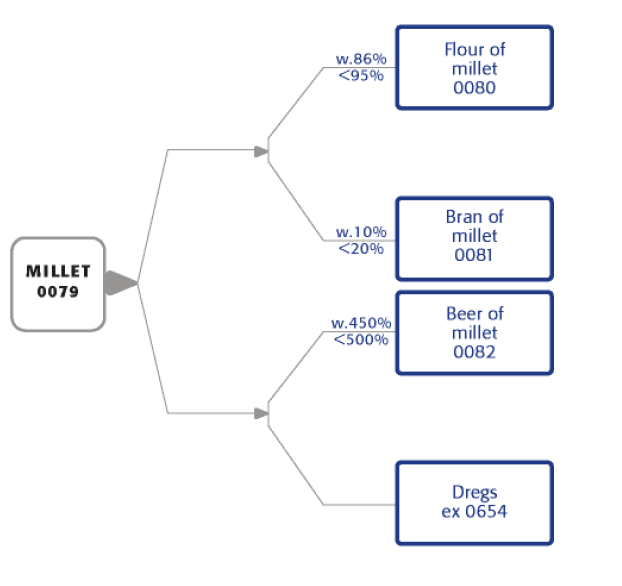
Group 391: Wastes from food and tobacco industry

Class 3912: Bran and other residues from the working of cereals or legumes

Subclass 39120: Bran and other residues from the working of cereals or legumes

FAO Expansion **39120.07: Bran of millet**

Figure 8: Commodity tree of millet in CPC



Flour of millet

23120.05

**Millet 0118**

Bran of millet 39120.07

Beer of millet 24310.03

*Dregs*

# **Appendix 1 Examples of solutions adopted to convert FAOSTAT data on agricultural commodities from FCL to CPC format**

In **one-to-one** cases old data are transferred to the new classification i.e. codes and definitions are re-assigned according to the new classification while data remain the same (Example 1).

Example 1:

data conversion from FCL to CPC in case of one-to-one type of link

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FCL | | | FCL🡪 CPC  conversion factor | CPC ver. 2.1 | | |
| code | descriptor | data (old format)  production quantity | code | descriptor | data (new format)  production quantity |
| 0125 | cassava | **4 082 903 tonnes** | **1** | 01520 | cassava | **4 082 903 tonnes** |

*Data are taken as example and refer to the production of cassava in Cameroon, 2011 (source: FAOSTAT)*

Also for **many-to-one** cases data conversion is straightforward as data in FCL are aggregated into CPC. Such an aggregation entails a loss of information, as the target classification is less detailed than the source one (Example 2).

Example 2:

data conversion from FCL to CPC in case of many-to-one type of link

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FCL | | | FCL 🡪 CPC  conversion factor | CPC ver.2.1 | | |
| code | descriptor | data (old format)  production quantity | code | descriptor | data (new format)  production quantity |
| 0430 | okra | **5 784 000 tonnes** | **Σ** | 01239 | other fruit bearing vegetables | 5 784 000 +  27 557 000=  **33 341 000 tonnes** |
| 0463 | other vegetables | **27 557 000 tonnes** |

*Data refer to the production of okra and other fresh vegetables in India, 2011 (source: FAOSTAT)*

Not to lose information in FAOSTAT, many-to-one cases are turned into one-to-one correlations: first the target classification is expanded further according to the detail available in FCL (new CPC expanded codes 01239.01 and 01239.90 in Example 3) and then the “key method” is applied as in Example 1. When detail in CPC 2.1 expanded is not yet sufficient, the classification is expanded further for FAOSTAT purpose.

Example 3:

FCL - CPC data conversion when many-to-one are turned into a one-to-one correlations (codes in **bold blue** text are the CPC expanded codes)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FCL | | | FCL 🡪 CPC  conversion factor | CPC ver.2.1 expanded | | |
| code | descriptor | data (old format)  production quantity | code | descriptor | data (new format)  production quantity |
| n/a | n/a | n/a |  | 01239 | other fruit-bearing vegetables | 33 341 000 tonnes |
| 0430 | okra | **5 784 000 tonnes** | **1** | **01239.01** | okra | **5 784 000 tonnes** |
| 0463 | other vegetables | **27 557 000 tonnes** | **1** | **01239.90** | other fruit-bearing vegetables n.e.c. | **27 557 000 tonnes** |

*Data refer to the production of okra and other fresh vegetables in India, 2011 (source: FAOSTAT)*

More difficulties are faced for one-to-many and many-to-many types of links. In these cases data have been converted based on statisticians’ best judgment according to the *dominant* correspondence. Coefficients of conversion have not been calculated, given the lack of information in both formats for at least one year and, therefore, the risk to threaten data quality in the conversion. The conversion keys assigned are 1 and 0 exclusively.

**One-to-many** relations between FCL and CPC mainly concern agricultural (primary) vs. industrial (processed) products. For example, fresh and dried fruit in FCL are sometimes classified together while they are separated in CPC. This is due to the fact that CPC is closely linked to the International Standard Industrial Classification of All Economic Activities (ISIC) and dried fruit is considered as an output of the manufacturing industry and not of agriculture. The solution adopted for data conversion in FAOSTAT when dried fruit is not dedicated a specific class (as in the case of dates) is to associate FCL data only to the items in the agricultural section of CPC, leaving blanks in correspondence of the industrial goods section. In Example 4 below, the one-to-many correlation is converted into one-to-one, assigning the conversion factor “1” to the class that, based on statistician’s best judgment, is the one better covering the FCL boundaries (dominant correspondence). In the metadata it will be noted that 01314 may, in some years for some countries, include information on dates dried on farm.

Example 4:

data conversion from FCL to CPC in case of one-to-many type of link

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FCL | | | FCL 🡪 CPC  conversion factor | CPC ver.2.1 expanded | | |
| code | descriptor | data (old format)  production quantity | code | descriptor | data (new format)  production quantity |
| 0577 | dates (fresh+dried) | **724 894 tonnes** | **1** | 01314 (agriculture) | dates, fresh | **724 894 tonnes** |
| **0** | 214190.03 (industrial) | dates, dried | **0** |

*Data refer to the production of dates in Algeria, 2011 (source: FAOSTAT)*

In **many-to-many** cases, which represent a minority of cases in the FCL-CPC correlations, CPC is modified and aligned to FCL.

In Example 5, the FCL code 0619 put “subtropical fruit” together with “fruit fresh n.e.s.” while in CPC subtropical fruit is classified with “other tropical and subtropical fruits, n.e.s.” (01319). This generates a mismatch between the two classifications. Given the impossibility to estimate split ratios, and not to introduce breaks in the series, CPC is adapted and aligned to FCL (Example 6): the component “subtropical fruit” in CPC is moved under “other fruits n.e.c.” as in FAOSTAT (01359.90). Definitions in the metadata are adjusted accordingly.

Example 5:

many-to-many correlations between FCL and CPC concerning tropical, subtropical and other fruit n.e.c.

|  |  |  |  |
| --- | --- | --- | --- |
| FCL | | CPC ver.2.1 expanded | |
| FCL code | FCL descriptor | CPC code | CPC descriptor |
| 0603 | fruit tropical fresh, n.e.s. | 01319 | other tropical and subtropical fruit, n.e.c. |
| 0619 | fruit fresh, n.e.s. (incl. subtropical) |
| 01359.90 | other fruits, n.e.c. |

Example 6:

data conversion from FCL to CPC in case of many-to-many type of link

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| FCL | | | FCL 🡪 CPC  conversion factor | CPC ver.2.1 expanded | | |
| code | descriptor | data (old format)  production quantity | code | descriptor | data (new format)  production quantity |
| 0603 | fruit tropical fresh, n.e.s. | **52 684 tonnes** | **1** | 01319 | other tropical **and subtropical** fruit, n.e.c. 🡪 other tropical fruit, n.e.c. (excluding subtropical fruit) | **52 684 tonnes** |
| 0619 | fruit fresh, n.e.s. (incl. subtropical) | **193 686(E) tonnes** | **1** | 01359.90 | other fruit, n.e.c. 🡪 other fruit, n.e.c. (**including** subtropical fruit) | **193 686(E) tonnes** |

*Data refer to the production of tropical fruit n.e.s. and fruit n.e.s. in Ecuador, 2011 (source: FAOSTAT; (E) = FAO estimates)*

1. The FCL structure and definitions are available on FAO Statistics Division at: [www.fao.org/waicent/faoinfo/economic/faodef/faodefe.htm](http://www.fao.org/waicent/faoinfo/economic/faodef/faodefe.htm) [↑](#footnote-ref-1)
2. <http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=28>; <http://unstats.un.org/unsd/publication/SeriesM/SeriesM_34rev4e.pdf> [↑](#footnote-ref-2)
3. Products follow the SNA definition i.e. *all output of economic activities* that can be the object of domestic or international transactions or that can be entered into stocks including transportable goods, non transportable goods, services and other products. [↑](#footnote-ref-3)
4. Valentina Ramaschiello, “*CPC Ver.2 Review and Harmonization with Food and Agriculture Statistics in FAO*”. Food and Agriculture Organization of the United Nations. presented at the Expert Group Meeting on International Classifications, UNSD, New York, May 2011, http://unstats.un.org/UNSD/class/intercop/expertgroup/2011/AC234-15.PDF [↑](#footnote-ref-4)
5. A classification at the lowest aggregation level is directly recoded to the revised classification. For example, the old code 12345 is recoded to 56789 and the historical data for 12345 are assigned to 56789. This method, also called “*key method*”, assures a straightforward relationship between the old and the new results, as the old data are simply transferred to the new classification. The process and outcomes should, however, be documented and communicated to the users. The “key method” is described in Gert Buiten, Jarl Kampen and Sidney Vergouw, 2009, “*Producing historical time series for STS-statistics in NACE Rev.2*”, Discussion paper (09001), Statistics Netherlands. <http://www.cbs.nl/NR/rdonlyres/A8A9AB3B-37F6-480A-BA76-253979DED22D/0/200901x10pub.pdf> [↑](#footnote-ref-5)
6. FAO. 2011. “*Food Balance Sheet: A Handbook*” <http://www.fao.org/docrep/003/X9892E/X9892E00.HTM> [↑](#footnote-ref-6)
7. Andrew Hancock, “*Best Practice Guidelines for Developing International Statistical Classifications*”, presented at the Meeting of the Expert Group on International Classifications, UNSD, New York, May 2013 <http://unstats.un.org/unsd/class/intercop/expertgroup/2013/AC267-5.PDF> [↑](#footnote-ref-7)
8. FAO. 2011. “*Food Balance Sheet: A Handbook*” <http://www.fao.org/docrep/003/X9892E/X9892E00.HTM> [↑](#footnote-ref-8)