Dealing with commodities in Navigocorpus. Offering tools and flexibility

Jean-Pierre Dedieu, Silvia Marzagalli

Revue de l'OFCE

140

EIGHTEENTH-CENTURY INTERNATIONAL TRADE STATISTICS

SOURCES AND METHODS

edited by Loïc Charles and Guillaume Daudin

ofce

edited by Loïc Charles and Guillaume Daudin

EIGHTEENTH-CENTURY INTERNATIONAL TRADE STATISTICS
SOURCES AND METHODS

ofce

Revue de l'OFCE

OFCE

L'Observatoire français des conjonctures économiques est un organisme indépendant de prévision, de recherche et d'évaluation des politiques publiques. Créé par une convention passée entre l'État et la Fondation nationale des sciences politiques approuvée par le décret n° 81.175 du 11 février 1981, l'OFCE regroupe plus de 40 chercheurs (es) français et étrangers. « Mettre au service du débat public en économie les fruits de la rigueur scientifique et de l'indépendance universitaire », telle est la mission que l'OFCE remplit en conduisant des travaux théoriques et empiriques, en participant aux réseaux scientifiques internationaux, en assurant une présence régulière dans les médias et en coopérant étroitement avec les pouvoirs publics français et européens. Philippe Weil a présidé l'OFCE de 2011 à 2013, à la suite de Jean-Paul Fitoussi, qui a succédé en 1989 au fondateur de l'OFCE, Jean-Marcel Jeanneney. Depuis 2014, Xavier Ragot préside l'OFCE. Il est assisté d'un conseil scientifique qui délibère sur l'orientation de ses travaux et l'utilisation des moyens.

Contact
OFCE, 69 quai d'Orsay 75340 Paris cedex 07
Tel. : +33(0)1 44 18 54 87
mail : revue@ofce.sciences-po.fr
web : www.ofce.sciences-po.fr

Dépôt légal : juillet 2015 ISBN : 978-2-312-03378-5
N° ISSN 1265-9576 – ISSN en ligne 1777-5647 – © OFCE 2015

Contents

EIGHTEENTH-CENTURY INTERNATIONAL TRADE STATISTICS

edited by Loïc Charles and Guillaume Daudin

The opinions expressed by the authors are their own and do not necessarily reflect the views or positions of the institutions to which they belong.

DEALING WITH COMMODITIES IN NAVIGOCORPUS
OFFERING TOOLS AND FLEXIBILITY

Jean-Pierre Dedieu
Framespa (CNRS – Université Jean Jaurès, Toulouse)
IAO (CNRS – Université de Lyon)

Silvia Marzagalli
Centre de la Méditerranée Moderne et Contemporaine, Nice
Institut Universitaire de France

From 2008 to 2011, we created Navigocorpus, an online data-base on shipping. While conceiving the data-base structure, we aimed at preserving data as close as possible to the way they appear in the sources – spelling and languages included – but also at developing a series of tools to handle the mass of data stored in the database. This paper deals with the way we processed cargoes and explains the three possibilities we offer to users. First, they can query a field containing a standardized English translation of cargo items. Secondly, they can create their own categories of classification in a on-the-way coding field, according to the specific needs of their research. Finally, they can query a permanent coding which provides, though a codified string of characters, information on the raw material, elaboration process and use of the product. A few concrete examples illustrate these features.

Keywords: eighteenth century, navigation statistics, Europe, economic history, commodity classification, database, standardization.

e-mail: jean-pierre.dedieu@ens-lyon.fr; silvia.marzagalli@unice.fr

From 2008 to 2011, the French Agence Nationale de la Recherche sponsored a research project called Navigocorpus.[1] The program aimed at creating a common database on shipping to cope with different kinds of sources related to ship voyages and to provide, both through the density of information and through the facilities it offers to researchers, a useful online tool for new research in maritime history. The database was designed to allow the integration and preservation of existing databases created by individual researchers and the collection of new information.[2]

The general philosophy of the project relies on two different and apparently contradictory goals: on the one hand, we provide data as close as possible to the way they appear in the sources; on the other, we developed a series of links and fields which transform the database into something radically different from a collection of loose data. For instance, we provide ships and captains with an identifier, so that it is possible to track them easily through different sources. Finally, we developed a series of tools to handle the mass of data stored in the database.

Navigocorpus at present (October 2014) contains information on over 75,000 ship voyages in the eighteenth and nineteenth centuries. Creating the database has been a challenge for two main reasons: first, contrary to most database projects, Navigocorpus was not conceived or structured to answer a specific research question or set of questions. We were therefore obliged to keep the structure as flexible as possible so that researchers might easily retrieve and

1. Navigocorpus ("Corpus des itinéraires des navires de commerce, XVIIe-XIXe siècles" - "Database on the Itineraries of Merchant Ships, 17th-19th Centuries") is a research project sponsored by the French Agence Nationale de la Recherche (ANR-07-CORP-028) and coordinated by Silvia Marzagalli (Centre de la Méditerranée Moderne et Contemporaine, Nice) in collaboration with Jean-Pierre Dedieu (at that time, Laboratoire de Recherche Historique Rhône-Alpes, Lyon) and Pierrick Pourchasse (Centre de Recherche Bretonne et Celtique, Brest). Werner Scheltjens worked on this program as a post-doctoral fellow from 2008 to 2010. The database is available online at http://navigocorpus.org/. More on this project at http://navigocorpus.hypotheses.org/?lang=fr_FR.

2. The main theoretical problems we faced and the solutions we elaborated have been presented in a joint paper: Jean-Pierre Dedieu, Silvia Marzagalli, Pierrick Pourchasse and Werner Scheltjens, "Navigocorpus, a database for shipping information. A methodological and technical introduction", *International Journal of Maritime History*, XXIII, No 2, December 2011, p. 241-262. See also, by the same authors, "Navigocorpus at Work. A Brief Overview of the Potentialities of a Database", *International Journal of Maritime History*, XXIV, No 1, June 2012, p. 331-359.

handle the data they need for their specific research goals. Second, Navigocorpus is an open database in which new data can be added at any moment. In order to go beyond the simple digitalization of sources and to make their content manageable to researchers, we had to develop fields and tools for selecting and handling data according to their needs. The result is a powerful engine for research in maritime history, although its complexity requires that one has a good understanding of the database structure before he/she be able to extract information efficiently.

We believe that some of the solutions we adopted might be useful to other research projects facing similar issues. This paper deals specifically with the way we processed cargoes. Colleagues dealing with merchant accounts or balances of trade statistics face the same challenge as maritime historians: how should they best make hundreds of different terms manageable? Navigocorpus provides researchers with a considerable amount of data on cargoes and the circulation of products in time and space. At the moment, the database contains 98,000 records, written with 5,000 different spellings, related to commodities. Data have been collected in five languages (Danish, English, French, Italian and Spanish). We present here the general features of the database and the choices we made in order to process cargo-related data. We will take a few examples – notably in the category "fish", which illustrates the complexity of dealing even with apparently simple categories.

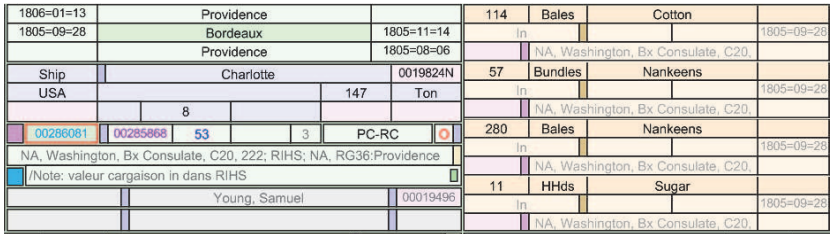
Before going into detail about how we handled cargoes, it is useful to understand that Navigcorpus is structured around the points the ship touched during its journey. A point is defined as a place (which we geo-referenced) through which a ship sailed, characterized by a date and an action (entering, clearing, passing through, etc.). The points are chronologically ordered and stored in a specific table. Information on the ship, captain, crew, cargoes, taxes, etc., are linked to the different points to which they pertain.

Most sources we inserted into Navigocorpus provide some information on the cargoes. We created a cargo-table in the database and inserted cargo items as individual records linked to the point to which information is related: a cargo entering the port of Marseille on a ship coming from Smyrna is linked to Marseille. We defined an item as a product stated by the source and expressed with the same unit of conditioning, weight or price. A ship cargo

listed by the source therefore originates as many records in the cargo table as the combination of items/units of measure. The 57 bundles and 280 bales of nankeens entering Bordeaux on 28 September 1805 on the ship *Charlotte* from Providence, for instance, generate two records in the cargo-table: nankeens, 57 bundles; and nankeens, 280 bales. Both are linked to the point Bordeaux, characterized by the date of entry and the action – in this case "IN", as the cargo items are being introduced into Bordeaux (Figure 1).

Figure 1. Example of cargo items on the *Charlotte*

a. Database main layout (cargo items on the right)



b. Cargo-table main layout

1805=09=28!	Cotton		Bordeaux	A0180923	In	Providence	B2837542	114	Bales
1805=09=28!	Nankeens		Bordeaux	A0180923	In	Providence	B2837542	57	Bundles
1805=09=28!	Nankeens		Bordeaux	A0180923	In	Providence	B2837542	280	Bales
1805=09=28!	Sugar		Bordeaux	A0180923	In	Providence	B2837542	11	HHds
1805=09=28!	Sugar		Bordeaux	A0180923	In	Providence	B2837542	68	Boxes
1805=09=28!	Tobacco		Bordeaux	A0180923	In	Providence	B2837542	66	Hhds

In constructing the cargo-table, we adopted the terminology provided by the sources without trying to impose any standardized notion of commodity. We believe that commodities do not exist as such and that they are the product of decisions taken by the actors:

> *In international trade we define commodities as materials and articles movable and procurable. The term "commodities" may be applied to the finest detail or to homogeneous groupings of the detailed classifications. There is therefore really no firm answer as to whether "fabrics" or "cotton fabrics" are "commodities"*[...]
>
> *The answer depends on the use made of the classification and thus depends on whether the products are sufficiently homogeneous in price, economic use, market, etc., to fit the needs of classification.*[3]

3. V. S. Kolesnikoff, "Commodity classification", in R. G. D. Allen and J. Edward Ely, *International Trade Statistics*, London, John Wiley and Sons, Inc., 1953, p. 50-81, here p. 55-56.

Sources, however, are not always explicit enough for the needs of researchers. Depending on the goals authorities had when collecting data on ships and cargoes, information on cargoes are more or less exhaustive. The lack of information introduces categories of commodities which artificially overlap with other categories provided in other entries of the database. Differences can show up even within the same sources. The American consul in Bordeaux generally clearly distinguished the nature of the cod imported by incoming United States ships. He stated, for instance, that the brig *Retrieve*, which entered Bordeaux on 1 July 1795, introduced 90 quintals of dry codfish and 1,300 quintals of green codfish; on 12 December 1797, however, the cargo of the schooner *Sally* is laconically described as "a parcel of cod". These two documentary units therefore provide three different commodities: cod, green codfish and dry codfish.[4] Similarly, with the general category "drygoods", some officers noted without further detail the manufactured goods put onboard whereas merchants and captains would record each item. For example, the outbound cargo of the *Eliza* of Salem, which sailed from Bordeaux on 15 October 1794, comprised "91½ tons of old claret; 100 pieces best Bordeaux brandy coloured & prepared, being fortified with 500 velts best 3/5th spirits 5 velts to each piece; 288 best silk stockings; 348 pairs Paris & other [stockings] – including woman & men, with work'd clocks, coulored, etc.; 1 piece superfine Brussels lace; 1 piece broader Brussels lace; 1 piece Valenciennes Cambric thick; 2 pieces Valenciennes Cambric finer; 2 pieces clear Cambric Handkerchief," etc., which the United States consul at Bordeaux systematically and laconically recorded as "wine, brandy and drygoods".[5]

Whenever general categories show up, researchers have to make a choice. A scholar interested in salt fish will have to decide whether he or she takes the unspecified cod into account. His or her specific knowledge will probably induce him or her to decide

4. National Archives and Record Administration, RG 84, Bordeaux Consulate, 26, Consular register of vessels entered and cleared, 1795-1797.

5. Peabody Essex Museum, Salem (Mass.), Derby family papers, Vol. 4. Elias Haskett Derby Invoice book, Inward cargoes, 1786-1796, No. 100: Invoice of the sundry merchandises shipped on the American Vessels the Eliza of & for Salem, Stephen Phillips master, on order of said Phillips for account & risk of M. Elias H. Derby of said place. See Silvia Marzagalli, *Bordeaux et les États-Unis, 1776 – 1815 : politique et stratégies négociantes dans la genèse d'un réseau commercial*, Geneva, Droz, 2015, p. 225.

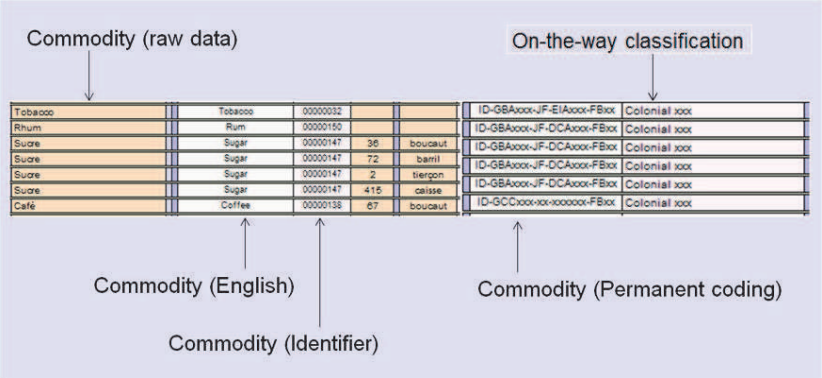
that "cod" entering Barcelona in 1790 is salt cod, whereas "cod" entering Le Havre the same year is green cod[6]. Other criteria will have to be taken into account to decide that the *Sally* probably had dry fish on board: the schooner arrived directly from Newfoundland, and the fact that she entered Bordeaux in December makes it likely that she had spent the summer on the shores to dry the fish. But we cannot make decisions of this kind throughout the database, both for a practical reason – we do not have universal knowledge on all trades and commodities – and for a methodological reason – we do not want to impose our categories on researchers. If we conceive it proper to classify in terms of salt/fresh fish, researchers might think of other, potentially unlimited categories for conceptualizing and reducing the enormous variety of commodities provided by the sources in a more manageable way.

In order to allow the greatest flexibility to Navigocorpus users to meet their own needs, we created the ability to classify "on-the-way" (or "on-the-run"): we created a field in which the researcher can add a term or a string of terms which correspond to the classification categories he or she wishes to adopt. In the example in Figure 2, we imagined that a researcher working on late eighteenth-century trade chose to label as "colonial goods", rum, sugar, coffee and tobacco. Working on fish consumption in early modern Catholic Europe can similarly induce a researcher to label whales as "fish", as contemporaries considered them to be. Classifying a whale as a mammal would be completely inadequate for the purpose, but might fit for someone working on the twenty-first century. Other researchers might want to create a category of "protected animals", and so on. This on-the-way coding is fast and temporary but highly unstable, as it is related to specific research, to a specific context and to a personal classification.

However, we did not want users to have to face the task of dealing with thousands of different cargo items on their own. One of the most obvious problems in a database such as Navigocorpus comes from the variety of languages and spellings used in the sources to designate products. There was an evident need to help users wishing to locate an item whatever the language of the

6. On the areas of consumption of green and salt cod, see *Atlas historique du Canada*, vol. I : *Des origines à 1800*, Montréal, 1987.

Figure 2. Dealing with commodities in Navigocorpus
From sources to standardization through flexibility



Raw data: spelling as in the sources
Commodity (English): standardized equivalent in English (and identifier)
Permanent coding: alphanumeric string added by the Navigocorpus team to data
"On the way" classification: applied by users for the scope of their research.

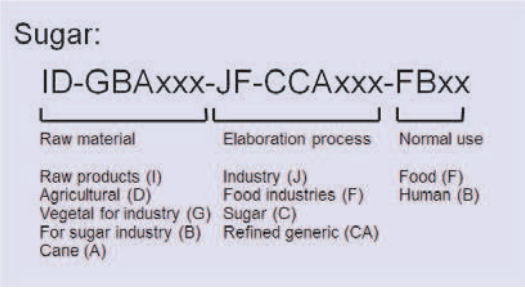
sources and the spelling. Looking for trade in anchovies, the user obviously wants to retrieve in one click information about *acciughe*, *anchois* and *anchovies*. Navigocorpus handled this aspect by introducing an additional field in which different spellings and languages are standardized in English. The user can query the database in any language used in Navigocorpus and access all the related corresponding languages, although the original source-based definition of the item is preserved, so that it is, for instance, possible to correlate the use of a specific terminology to study institutionalized categories, linguistic aspects, or to point to differences among commodities that were self-evident for contemporaries but that we might overlook (for instance, different terms to designate the same fish in the same port at the same time according to the treatment it underwent for preservation). We also provided each entry of a standardized commodity with a numeric identifier. In so doing, we kept the distinctions in the sources: *acciughe*, *anchois* and *anchovies* are provided with the same identifier, but *acciughe salate, anchois salées* and *salt anchovies* are provided with another identifier.

We therefore have maintained separate categories but also provided tools to users so they can save time while retrieving information from the database without having to query individually, one after the other, *salt anchovies*, salt cod, stockfish, etc., when

trying to locate salt fish. For that purpose, in order to control the instability of the data, we added a third level of codification through permanent coding. The latter is composed of a string of characters to univocally identify the commodity in question, whatever the name given by the sources. This phase involves a most difficult task of standardization, which kept generations of economists, statisticians and customs officers busy in largely fruitless efforts. The fact that we do not materially manage the product, as we are dependent on the partial descriptions given by the sources, made the task even more delicate. Moreover, we had to meet the very special and unpredictable needs of researchers, not the clear-cut requirements of an administration.

Practitioners agree that classifications of commodities rest on one of the following three criteria: the raw material; the elaboration process; or the most frequent use of the product.[7] Accordingly, a wine bottle can be classified as a mineral (glass); or as an industrial product; or as a (food) container. In the pencil-and-paper era, people who classified items had to choose one of these criteria and stick to it, making only slight and partial allowance for the other two when absolutely necessary to avoid making the result hopelessly unmanageable. Computers removed these constraints, as they can easily retrieve complex queries from complex character strings. We therefore could adopt a multi-criteria classification. Our classification includes the three standard criteria based on material, elaboration and use (Figure 3).

Figure 3. Composition of permanent coding. The example of eighteenth-century white sugar imported from Saint-Domingue



7. Kolesnikoff, "Commodity classification", p. 50–81.

The first part of the code provides information on the raw material; the second part provides additional information on the elaboration processes; the last part helps identify the use of the product. The meaning of each coding constitutive unit (letter) depends on its position. The length and number of elements composing each part is always the same in order to facilitate queries, with lower-case "x" filling void (non-significant) positions. The highly structured, context-dependent and modular nature of the coding sequence makes a correct interpretation by means of interpretative tables a relatively easy task (see Appendix 1 for more details).

Such permanent coding responds to different needs. First, it sticks as closely as possible to an effective description of the product, taking into account every factor which might affect its insertion into economic, social and cultural contexts.[8] Second, it provides three possible entries to researchers, depending on the interest they have. Third, we provide a grammar which can be expanded to cope with unexpected cases, as well as with partially documented ones: generic designations will be coded using the first positions only; more precise ones will use the same root and will be completed by extra characters. Finally, the hierarchically ordered sequence of coding elements, fully in consonance with the way computers work, allows a natural grouping of commodities into classes.

The elaboration of such codes requires complete knowledge of the products. It is therefore a considerably complex and time-consuming task for which scientific collaboration among research groups with different expertise would be very beneficial. At an early stage, operators could process the commodity, even when precisely described, by using the first, more generic part of the code. In a later stage, they might add to the coding whenever extra information on elaboration processes, uses and raw materials made that possible. In all instances, full coding implies full research on products based on contemporary technical dictionaries, from

8. The last sequence, concerning use of the product, may change when coding a commodity designated by the same name depending on the social function it is supposed to fulfill in the context in which it appears: imported lemon juice to Marseille in the eighteenth century did not have the same use if in bottles (for drinking) or casks (for industrial use), whereas it can be classified as a drug if delivered, for instance, to the Navy or to the East India Company.

d'Alembert's *Encyclopedia*[9] to less known but nonetheless highly useful works such as the *Technological listing...* published at the beginning of the twentieth century by the French statistical office, which is a fairly complete description of products and elaboration processes.[10]

We believe that the coding procedure we suggest constitutes a useful collaborative tool for storing and making available specialized knowledge usually restricted to expert circles. The permanent coding obviously cannot be freely tampered with, but its use is optional: the on-the-way coding provides Navigocorpus users with the margin of liberty they need.

Conclusion

In handling commodities within Navigocorpus, we have tried to overcome constraints and methodological challenges. We wanted to preserve data as they appear in the sources, and we managed to do that. We also wanted to do something more than providing users with an enormous amount of rough data, so we provided a standardized English translation of items. We invented a multi-criteria classification which users are free to use or not – knowing that at this stage we have not fully processed all cargo data we collected. And through on-the-way coding we allowed for the greatest possible flexibility in adding classifications that users think proper according to their research goals.

9. An excellent online version of the same by the ARTFL Encyclopedie Project of the University of Chicago (http://encyclopedie.uchicago.edu/) (consulted: May 2014).
10. Ministère du Travail et de la Prévoyance Sociale. Statistique générale de la France, *Répertoire technologique des noms d'industries et de professions français – anglais – allemands avec notices descriptives sommaires, suivi des trois listes alphabétiques des noms allemands, anglais et français*, Paris/Nancy, Berger-Levrault et Cie, 1909, 771p.

Appendix:
Commodities permanent coding adopted by Navigocorpus
The example of fish

Each commodity is described by a permanent coding string of the form:

xx-xxxxxx-xx-xxxxxx-xxxx

Permanent coding is composed of three legs. The first two are in turn composed of two parts. The first leg describes raw materials. The second describes the elaboration process. When the commodity is mere raw material, this second leg remains empty (empty meaning a series of xxxxx). The third leg expresses the use generally made of the commodity (human food, fodder, industrial process, etc.).

Each position of the coding string must be materially filled with a letter, as positions are significant. Positions which should remain blank must therefore be filled with "x". Hyphens have no meaning: their function is to make the string easier to read. Segments match meaning in the following way:

-[1] [2] Material, basic product
. Class
. Material

- [3] [4] Elaboration process
. Class
. Special description

- [5] General use given to the commodity

xx-xxxxxx-xx-xxxxxx-xxxx
1 2 3 4 5

Dictionary

We present here extracts from the general dictionary for permanent coding we are presently elaborating.

a) First leg: raw material

Ix: Raw products

IA: raw fishery products

IA-FAxxxx: fish in general

IA-FAAxxx: fish, without further detail

IA-FBxxxx: great fishery products

IA-FBBAxx: whale
IA-FBBABx: sperm whale
IA-FBBCxx: cod
IA-FBBExx: seal
IA-FCxxxx: fishery of open sea products
IA-FCAxxx: products of high sea fishery (generic)
IA-FCABxx: fish salted on board (generic)
IA-FCBxxx: fish from high sea fishery (special items)
IA-FCBAxx: herring
IA-FCBBxx: pilchard
IA-FCBCxx: mackerel
IA-FCBDxx: anchovy
IA-FCBExx: tuna
IA-FCBEBx: blue fin tuna
IA-FDxxxx: inshore fishery, coastal fishing products
IA-FDAxxx: fish from coastal fishing (general)
IA-FDAAxx: fresh fish
IA-FDABxx: skate
IA-FDBxxx: shellfish products
IA-FDBAxx: salvage shellfish products
IA-FDBAAx: oysters
IA-FDBABx: mussels
IA-FDBBxx: shell culture products
IA-FDBBAx: oysters
IA-FDBBBx: mussels
IA-FDBCxx: shell (undetermined breed)
IA-FDBCAx: oysters
IA-FDBCBx: mussels
IA-FFxxxx: river fishery
IA-FFAxxx: generic river fish
IA-FFBxxx: special fishing
IA-FFBBxx: salmon

b) Second leg: elaboration processes

JF-Hxxxxx: preserves
JF-HAxxxx: preserves (generic)
JF-HBxxxx: preserves through additives
JF-HBAxxx: preserves through additives (generic)
JF-HBBxxx: sugared preserves
JF-HBBAxx: sugar preserved (generic)
JF-HBBBxx: candied (unused)
JF-HBBCxx: industrial jams
JF-HBBDxx: syrups

JF-HBCxxx: salted preserves
JF-HBCAxx: salted preserves (generic)
JF-HBCBxx: ham
JF-HBCCxx: [code left empty for possible future use]
JF-HBCDxx: bacon
JF-HBDxxx: [code left empty for possible future use]
JF-HBExxx: oil preserves
JF-HBFxxx: marinated preserves
JF-HBGxxx: pickles
JF-HCxxxx: dried preserves
JF-HDxxxx: mixed dried and salted preserves
JF-HExxxx: [code left empty for possible future use]
JF-HFxxxx: chilled preserves
JF-HFBxxx: chilled
JF-HFCxxx: iced
JF-HFFxxx: frozen
JF-HGxxxx: paste preserve, pressed preserves
JF-HGBxxx: meat pastes
JF-HHxxxx: canned preserves
JF-HIxxxx: vinegar and derived condiments
JF-HIAxxx: vinegar
JF-HIBxxx: mustard
JF-HKxxxx: smoked preserves
JF-HKAxxx: smoked preserves in general

c) *Third leg: uses given to the commodity*

Fxxx: food
FAxx: food (undetermined)
FBxx: human food
FBAx: basic product
FBBx: luxury products
FBCx:
FBDx: drinking products
FBEx:
FBFx: seasonings
FBGx: fats
FDxx: mixed human and animal food
FExx: animal food, fodder
FFxx: mixed human and animal food
FGxx: mixed food and industrial use

d) Examples of combinations of the three legs

IA-FCBAxx-JF-HAxxxx-FBxxx: preserved herring
IA-FAAxxx-JF-HBCAxx-FBxxx: salt fish
IA-FCBAxx-JF-HBCAxx-FBxxx: salt herring
IA-FFBBxx-JF-HBCAxx-FBxxx: salt salmon
IA-FCBExx- JF-HBCAxx-FBxxx: salt tuna
IA-FCBExx- JF-HBExxx-FBxxx: tuna preserved in oil
IA-FCBAxx-JF-HBFxxx-FBxxx: marinated herring
IA-FCBExx-JF-HBFxxx-FBxxx: marinated tuna