EFSA algorithm for mapping the FoodEx2 codes in scientific data from EU monitoring programmes on certain residues in food to the EFSA residues legal limits database

Summary

The present document shortly describes the algorithm developed by EFSA and applied to the FoodEx2 coding system in order to retrieve and match the food classification codes in the EFSA Legal Limit data base (LLDB) with the ones returned by the data provider in their national datasets stored in EFSA Scientific Data Warehouse (SDWH).

This algorithm is executed by EFSA during the data validation step, as part of the Business Rules (BR) engine, when inserting a dataset through the Data Collection Framework platform (DCF) in the frame of the Chemical Monitoring data collection (ChemMon) for the year 2020. Then, a validation of the single data records in the dataset against the corresponding MRL is performed.

Background

EFSA is mandated to collect analytical results generated by Member States in the frame of the control activities carried by EU national authorities to monitor residues of pesticide, veterinary medicinal products (vmpr) and other residues in food and feed samples of plant and animal origin[[1]](#footnote-2),[[2]](#footnote-3). In the context of the national results transmission to EFSA, EFSA has set-up a validation procedure based on BR, whose application is aimed at ensuring the data submitted through the DCF are fit for purpose for use in EFSA’s scientific assessments and at improving the quality of the data collected and stored in the SDWH. Some of the BR are based on the newly developed EFSA internal database on the EU MRL for pesticide and vmpr residues to validate the consistency and plausibility of the information reported by the data providers on e.g. the assessment of the compliancy of the analytical results against the legal maximum residue level of the analysed substances in the tested samples.

The above results on national monitoring activities are coded and stored in the DDWH according to SSD2 data model, which makes use of the FoodEx2 classification system to describe the matrix, the production method and the food product treatment of the sample tested. The FoodEx2 coding system is based on the ‘MTX’ catalogue.

The EFSA internal LLDB is build starting from the structure of the ‘food’ MRL classification system in place at EU level under the two legal frameworks on MRL for pesticides and vmpr residues, respectively; these two classification systems are different between themselves and do not match with the FoodEx2 classification system. Therefore, for EFSA to validate the coded national monitoring results against the MRL database, an automated procedure (algorithm) was needed to map the FoodEx2 codes returned with the transmission of the national monitoring results to the corresponding food codes associated to the MRL values in to the LLDB. This mapping algorithm is specifically created for the EFSA validation of the pesticide and vmpr residues monitoring data. The national data provider will not necessarily need to know the details of this mapping as it is automatically applied by EFSA. Therefore, its short description is briefly outlined in this document for documentation’s purpose but may be useful for the national data providers that wish implementing the algorithm in their local data repository systems as well.

All resources linked to this document (e.g. input tables, catalogues, business rules, schema definitions) are published separately in Knowledge Junction in machine readable format and human readable, where appropriate.

The algorithm description

Table 1 describes the configuration file used by the algorithm.

Table 1 – Configuration table ‘LLDB\_FOODCLASSIFIER\_CONFIG’

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column name | Data Type | Column reportability (Mandatory)(a) | Description | SSD2 Catalogue |
| Regulation | VARCHAR  (15 CHAR) | Y | Legislation – representative of the data domain, e.g. progLegalRef=Regulation 396/2005 for the pesticide residues domain. | LEGREF |
| IncludeFacets | VARCHAR2  (4000 CHAR) | N\* | List of facets (separated by ‘$’) required to be in the FoodEx2 (or their descendant in the associated hierarchy) in order to be mapped | MTX\*\* |
| ExcludeFacets | VARCHAR2  (4000 CHAR) | N\* | List of facets (separated by ‘$’, where F00 stands for BaseTerm) required NOT to be in the FoodEx2 (or their descendant in the associated hierarchy) in order to be mapped | MTX\*\* |
| FoodClass | VARCHAR  (15 CHAR) | Y | FoodClassification | <catalogueClass> |
| CatalogueClass | VARCHAR  (15 CHAR) | Y | Name of the catalogue applied |  |
| baseTerm\_hier | VARCHAR  (15 CHAR) | N | Name of the hierarchy of the MTX in which the base term mentioned in the IncludeFacets or ExcludeFacets are taken |  |
| Mapping\_Order | NUMBER  (10) | Y | Numerical order used by the algorithm |  |

(a): Y=Mandatory; N=Optional

\* Mandatory (i.e. ‘Y’) if the other column ‘ExcludeFacets’ or ’IncludeFacets’ is not provided

\*\* Repeatable field separated by ‘$’, base terms should be coded as F00.’MTXCODE’

The table LLDB\_FOODCLASSIFIER\_CONFIG (Table 1) will feed the algorithm.

The algorithm is constructed starting from this table and it operates according to the following steps:

1. Select the Domain you want to use by filtering the configuration table **(given in INPUT)** from the ‘Regulation’ column
2. Add to the FoodEx2 code **(given in INPUT)** the implicit facets (i.e. attribute ‘ALLFACETS’ in the ‘MTX’ catalogue):
   1. For non-repeatable facets, the code for the explicit one is kept (if any);
   2. For repeatable facets:
      1. If the explicit facet code is a child of an implicit facet code in the specific facet hierarchy, then replace the implicit facet code with the explicit one;
      2. If the explicit code is not a child of any implicit facet code, then add it to the list of facets codes.
3. Starting from the row with smaller ‘Mapping\_Order’ number:
   1. Get the children of ‘IncludeFacets’ and ‘ExcludeFacets’ codes under the specific facet hierarchy or the hierarchy expressed in the column ‘baseTerm\_hier’;
   2. Check if the resulting FoodEx2 code (see above in Step 2) has each term of the value of ‘IncludeFacets’ (or their children obtained in above Step 3.a);
   3. Check if the resulting FoodEx2 code (Step 2) does not have the complete list of facets of the value of ‘ExcludeFacets’ (of their children);
   4. Verify if the number of facets expressed for any category for the InlcudeFacets is less or equal to the ones in the FoodEx2 code (Step.2).
   5. Check if (b), (c) and (d) are verified
      1. TRUE: the algorithm ends and the value of ‘FoodClass’ is selected **(OUTPUT);**
      2. FALSE: repeat the step.3 considering the following row in the configuration file
4. If the table has been fully checked and no mapping code has been found, then the code expressed in the row identified by IncludeFacets=“Other” will be selected **(OUTPUT)**.

In Table 2 a possible example of a configuration file is reported

Table 2 – Examples of the configuration table of the mapping algorithm for the pesticide residues domain

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Regulation** | **IncludeFacets** | **ExcludeFacets** | **FoodClass** | **CatalogueClass** | **baseTerm\_hier** | **Mapping\_Order** |
| N027A | F00.ROOT |  | P1200000A | MATRIX | FEED | 1 |
| N027A | F00.A03QA |  | PX100004A | MATRIX | EXPO | 2 |
| N027A | F00.A03QB |  | PX100004A | MATRIX | EXPO | 3 |
| N027A | F00.A03QC |  | PX100004A | MATRIX | EXPO | 4 |
| N027A | F00.A03QD |  | PX100004A | MATRIX | EXPO | 5 |
| N027A | F27.ROOT$F27.ROOT |  | XXXXXXA | MATRIX |  | 6 |
| N027A | F21.A07RY |  | P1070000A | MATRIX |  | 7 |
| N027A | F27.A01DH |  | P0130010A | MATRIX |  | 8 |
| N027A | F27.A01QX | F20.A0F4V | P1012000A | MATRIX |  | 9 |
| N027A | F27.A01QX |  | P1012010B | MATRIX |  | 10 |
| N027A | other |  | XXXXXXA | MATRIX |  | 11 |

With reference to Table 2:

1. Since the 1st line has IncludeFacets=’F00.ROOT’, no ExcludeFacets and baseTerm\_hier=’FEED’, it means that any FoodEx2 reported with a baseTerm in the FEED hierarchy will be mapped into the MATRIX code ‘P1200000A’=’Feed’ for the pesticide residues domain (Regulation=‘N027A’=’Regulation 396/2005’).
2. The next four lines of the above table tell us that if the FoodEx2 code with a baseTerm being one of four codes A03QA, A03QB, A03QC or A03QD (all codes relate to ‘Baby food’ samples) will be mapped to ‘PX100004A’=’Infant formula’. Since none of the mentioned codes have children codes in the Exposure hierarchy of the MTX catalogue (FoodEx2), then no other code will be mapped to ‘PX100004A’.
3. The 6th line tells us that any FoodEx2 code with two different source-commodities (F27 facet) codes will be mapped to ‘XXXXXXA’=’Not in list’.
4. The 7th line tells us that any FoodEx2 code with reported F21.A07RY (A07RY=’ Wild or gathered or hunted’) will be mapped to ‘P1070000A’=’ Wild terrestrial vertebrate animals’.
5. The 8th line tells us that any FoodEx2 code with F27.A01DH will be mapped to P0130010A. Thus, if a given mapping is applied to one FoodEx2 ‘parent’ code, the same mapping is done for all its children codes as in the F27 Source-commodities hierarchy. In this example, the code A01DH (‘Apples and similar’) has children in the source-commodities hierarchy, hence also any FoodEx2 code with F27.A01DJ (‘Apples’), or F27.A01DK (‘Crab apples’), or F27.A0DXF (‘Tejocotes’) will be mapped to P0130010A=’Apples’.
6. The following two lines (9th and 10th) tell us that any foodEx2 code containing F27.A01QX (‘Cow, ox or bull fresh meat’) will be mapped to P1012000A= Tissues (bovine)’ if the facet F20.A0F4V (‘Excluding visible fat’) is not reported, while it is mapped to P1012010B=’Muscle (bovine)’if F20.A0F4V (‘Excluding visible fat’) is explicitly reported.
7. Finally, the 11th line means that any other FoodEx2 code not matching one of the previous conditions is mapped to ‘XXXXXXA’=’Not in list’.

In Table 3 some examples are reported, in which the FoodEx2 codes are selected solely for the purpose of describing the algorithm, i.e. no consistency between the baseTerm and the facets values is checked for these examples.

**Table** 3 - Examples on the order of FoodEx2 to MATRIX codes mapping by the algorithm

| **Input**  **FoodEx2 code** | **Output**  **MATRIX code** | **Comment** |
| --- | --- | --- |
| A0BHH | P1200000A | This code is a Feed |
| A0BHH#F27.A01DH$F27.A01QX | P1200000A | This code is a Feed and is the first element the algorithm checks. So, no check for the double F27 is done |
| A03QD#F21.A07RY | PX100004A | The mapping is applied at the baseTerm level before checking the F21 facet |
| A03QC#F27.A01DH$F27.A01QX | PX100004A | The mapping is applied at the baseTerm level before checking the F27 facet |
| A02EP#F21.A07RY | P1070000A | The mapping on the F21 facet is applied |
| A02EP#F21.A07RY$F27.A01DH | XXXXXXA | The base term has already a F27 facet code implicitly included in the selected FoodEx2 term (F27.A02EP). So, the double F27 facets are checked |
| A02EP#F21.A0CAG | XXXXXXA | Neither the baseTerm, nor the implicit facets or the explicit F21 is present in the Configuration file. Line with Mapping\_Order = 11 applies |
| A01QX | P1012000A | The code A01QX has the implicit facet F27.A01QX, moreover the facet F20.A0F4V is not present, so line with Mapping\_Order = 9 is respected |
| A01QX$F20.A0F4V | P1012010B | The code A01QX has the implicit facet F27.A01QX; moreover, the facet F20.A0F4V is present, so line with Mapping\_Order = 10 is respected |

Abbreviations

|  |  |
| --- | --- |
| BR | Business Rule |
| DCF | Data Collection Framework |
| EFSA | European Food Safety Authority |
| EU | European Union |
| FoodEx2 | EFSA’s food classification and description system |
| LLDB | EFSA Legal Limit Data Base |
| MRL | Maximum Residue Level |
| MTX | FoodEx2 Matrix catalogue |
| SDWH | EFSA Scientific Data Warehouse |
| SSD2 | Standard Sample Description (ver. 2) |
| vmpr | veterinary medicinal products residues |

1. Regulation (EC) No 396/2005 of the European Parliament and of the Council of 23 February 2005 on maximum residue levels of pesticides in or on food and feed of plant and animal origin and amending Council Directive 91/414/EEC, avaible at: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32005R0396> [↑](#footnote-ref-2)
2. Council Directive 96/23/EC of 29 April 1996 on measures to monitor certain substances and residues thereof in live animals and animal products and repealing Directives 85/358/EEC and 86/469/EEC and Decisions 89/187/EEC and 91/664/EEC <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A31996L0023> [↑](#footnote-ref-3)