# Towards a Unique Data Model for Chemical and Microbiological Food Information

Claudia Machado, Silvia Viegas, Luísa Oliveira (✉)
Departamento de Alimentação e Nutrição do Instituto Nacional de Saúde Doutor Ricardo Jorge, IP (INSA)
Av. Padre Cruz, 1649-016 Lisboa, Portugal
(✉) *luisa.oliveira@insa.min-saude.pt*

**Keywords**

Data Model; Food Information; EFSA; Microbiology; Chemistry

## Introduction:

The integration of food data from research, monitoring, control, epidemiology and other sources is crucial to improve food safety and public health. Consequently, INSA launched the Portuguese Food Information Resource Programme (PortFIR) in a partnership with GS1 Portugal CODIPOR to create national food expert networks and sustainable databases on food composition, consumption and chemical and microbiological contamination [1]. Presently, the PortFIR data model is being developed.

Existing data models on food information usually refer to either chemical substances or microorganisms. However, for food safety, particularly for risk-benefit evaluation, a unique data model to compile Chemical and Microbiological Food Information (CMFI) would be a huge step forward, regarding data standardization and optimization of resources.

## Purpose:

The aim of this work is to explore the possibility of creating a unique PortFIR data model for the compilation, management and use of CMFI compatible with EFSA's chemical and microbiological calls for data and for nutrition applications, namely the EuroFIR European Food Composition Data Bank.

## Method:

The work was developed in two steps:

1) identification of existing relevant data models;

2) listing and comparison of all specified attributes.

## Results:

The data models identified as relevant were: a) EFSA's Standard Sample Description for Food and Feed (SSD), for chemical contaminants [2], b) Zoonoses Data Collection (ZDC), for microbiological agents and food borne outbreaks (FBO), composed by two different guidance documents [3, 4] and c) CEN/TC 387 prEN_16104 Food Data – Data structure (CEN/TC 387) [5]. The reasons to choose these references were the need to report data to EFSA and to update Portuguese data in the EuroFIR European Food Composition Data Bank whose structure formed the basis of CEN/TC 387 work. All attributes in each data model were listed independently of their classes, correspondence among models was cross-referenced and controlled terminology requirement was identified, attributes corresponding to aggregated values calculated from raw data potentially managed in other databases are identified, as shown in Table 1.

Table 1 - Attributes of SSD, ZDC and CEN/TC 387 Data Models

| Attribute | SSD | ZDC | CEN/TC 387 |
|---|---|---|---|
| Accreditation procedure for the analytical method * | ✓ | ✓ | ✓ |
| Analytical method code * | ✓ | ✓ | ✓ |
| Analytical method text | ✓ | × | ✓ |
| Area of origin for fisheries or aquaculture activities code * | ✓ | × | ✓ |
| Area of processing * | ✓ | × | ✓ |
| Area of the origin of the product | ✓ | × | ✓ |
| Brand name | ✓ | × | ✓ |
| CC alpha | ✓ | × | ✓ |
| Comment of the result | ✓ | × | ✓ |
| Country of origin of the product * | ✓ | × | ✓ |
| Country of sampling * | ✓ | × | ✓ |
| Day of expiry | ✓ | × | ✓ |
| Day of sampling | ✓ | × | ✓ |
| EFSA product code * | ✓ | × | ✓ |
| Expression of result * | ✓ | × | × |
| FBO code | × | ✓ | × |
| FBO food vehicle * | × | ✓ | × |
| FBO place of origin of problem * | × | ✓ | × |
| Frequency of the sampling * | × | ✓ | × |
| History of the disease and/or infection (F) | × | ✓ | × |
| Laboratory | ✓ | ✓ | × |
| Laboratory country * | ✓ | × | × |
| Laboratory sub-sample code | ✓ | × | ✓ |
| Legal limit for the result | ✓ | × | ✓ |
| Local organization country * | ✓ | × | × |
| Lot size unit * | ✓ | ✓ | ✓ |
| Method of production | ✓ | × | ✓ |
| Month of expiry | ✓ | × | ✓ |
| Month of sampling | ✓ | × | ✓ |
| Nature of evidence linking FBO cases with food vehicle * | × | ✓ | × |
| Number of samples (composite) | ✓ | ✓ | × |
| Other preventive measures than vaccination in place (F) | × | ✓ | × |
| Parameter code * | ✓ | ✓ | ✓ |
| Percentage of fat in the original sample | ✓ | × | ✓ |
| Product code * | ✓ | ✓ | ✓ |
| Product full text description | ✓ | ✓ | ✓ |
| Programme legal reference | ✓ | × | × |
| Region (F) | × | ✓ | × |
| Resistance profile | × | ✓ | × |

| Attribute | SSD | ZDC | CEN/TC 387 |
|---|---|---|---|
| Action taken * | ✓ | ✓ | × |
| Analytical method reference code | ✓ | ✓ | ✓ |
| Animal species (F) | × | ✓ | × |
| Area of origin for fisheries or aquaculture activities text | ✓ | × | ✓ |
| Area of sampling * | ✓ | × | ✓ |
| Arrival laboratory date | × | ✓ | ✓ |
| Category of animals (F) | × | ✓ | × |
| CC beta | ✓ | × | ✓ |
| Control programmes (F) | × | ✓ | × |
| Country of processing * | ✓ | × | ✓ |
| Day of analysis | ✓ | × | ✓ |
| Day of production | ✓ | × | ✓ |
| EFSA parameter code * | ✓ | × | × |
| Evaluation of the result * | ✓ | ✓ | × |
| FBO causative agent * | × | ✓ | × |
| FBO contributory factors * | × | ✓ | × |
| FBO mixed (other agents) | × | ✓ | × |
| FBO setting * | × | ✓ | × |
| Geographical distribution and size distribution of the herds, flocks and holdings (F) | × | ✓ | × |
| Ingredients | ✓ | × | ✓ |
| Laboratory accreditation * | ✓ | ✓ | ✓ |
| Laboratory sample code | ✓ | × | ✓ |
| Language * | ✓ | × | ✓ |
| Local organization | ✓ | × | × |
| Lot size | ✓ | × | ✓ |
| Manufacturer | ✓ | × | ✓ |
| Month of analysis | ✓ | × | ✓ |
| Month of production | ✓ | × | ✓ |
| More information (FBO) | × | ✓ | × |
| Notification system in place (F) | × | ✓ | × |
| Origin of food vehicle * | × | ✓ | × |
| Packaging * | ✓ | × | ✓ |
| Parameter text | ✓ | ✓ | ✓ |
| Percentage of moisture in the original sample | ✓ | × | ✓ |
| Product comment | ✓ | × | ✓ |
| Product treatment * | ✓ | ✓ | ✓ |
| Programme type * | ✓ | ✓ | × |
| Reporting year (F) | × | ✓ | × |
| Result code | ✓ | × | × |

| Attribute | SSD | ZDC | CEN/TC 387 |
|---|---|---|---|
| Result LOD | ✓ | × | ✓ |
| Result qualitative value * | ✓ | × | ✓ |
| Result value | ✓ | × | ✓ |
| Result value uncertainty | ✓ | × | ✓ |
| Reult value recovery | ✓ | × | ✓ |
| Sample type * | × | ✓ | × |
| Sampling details | × | ✓ | × |
| Sampling point * | ✓ | ✓ | ✓ |
| Sampling strategy * | ✓ | ✓ | ✓ |
| Source of information | × | ✓ | × |
| Type of legal limit * | ✓ | × | ✓ |
| Type of parameter * | ✓ | × | × |
| Vaccination description (F) | × | ✓ | × |
| Year of expiry | ✓ | × | ✓ |
| Year of sampling | ✓ | × | ✓ |

| Attribute | SSD | ZDC | CEN/TC 387 |
|---|---|---|---|
| Result LOQ | ✓ | × | ✓ |
| Result unit | ✓ | × | ✓ |
| Result value corrected for recovery * | ✓ | × | ✓ |
| Result value uncertainty (Standard deviation ) | ✓ | × | ✓ |
| Sample category * | × | ✓ | × |
| Sample's storage | × | ✓ | ✓ |
| Sampling method * | ✓ | ✓ | × |
| Sampling programme code | ✓ | × | ✓ |
| Sampling unit | × | ✓ | × |
| Storage date | × | ✓ | ✓ |
| Type of FBO * | × | ✓ | × |
| Type of result * | ✓ | × | ✓ |
| Year of analysis | ✓ | × | ✓ |
| Year of production | ✓ | × | ✓ |

**Attributes of aggregated values obtained from original raw data managed in other databases**

| | | | |
|---|---|---|---|
| Host's age group (ZA) * | × | ✓ | × |
| Livestock number (live animals) (F) | × | ✓ | × |
| Number of herds or flocks (F) | × | ✓ | × |
| Number of isolates tested (ZA) | × | ✓ | × |
| Number of new positive herds (F) | × | ✓ | × |
| Number of positive herds depopulated (F) | × | ✓ | × |
| Number of zoonoses autochtone cases (F) | × | ✓ | × |
| Number of zoonoses imported cases | × | ✓ | × |
| Percentage (%) of new positive herds (herd incidence) (F) | × | ✓ | × |
| Percentage (%) of positive herds coverage (F) | × | ✓ | × |
| Total number of deaths in FBO | × | ✓ | × |
| Total number of human cases (FBO) | × | ✓ | × |
| Total units positive for species (ZA) | × | ✓ | × |
| Total units tested (ZA) | × | ✓ | × |

| | | | |
|---|---|---|---|
| Host's sex (ZA) * | × | ✓ | × |
| Number of herds checked (F) | × | ✓ | × |
| Number of holdings (F) | × | ✓ | × |
| Number of multiresistance isolates | × | ✓ | × |
| Number of positive herds (F) | × | ✓ | × |
| Number of slaughtered animals (F) | × | ✓ | × |
| Number of zoonoses cases | × | ✓ | × |
| Number of zoonoses occupational cases | × | ✓ | × |
| Percentage (%) of positive herds (period herd prevalence) (F) | × | ✓ | × |
| Percentage (%) of positive herds depopulated (F) | × | ✓ | × |
| Total number of hospitalisations | × | ✓ | × |
| Total units positive (ZA) | × | ✓ | × |
| Total units positive *spp.*, unsepecified (ZA) | × | ✓ | × |

*Legend:* ✓ - *Present* × - *Absent*

* – *controlled terminology* *(F) – Farm* *(FBO) – Food borne Outbreak* *(ZA) – Zoonotic Agent*

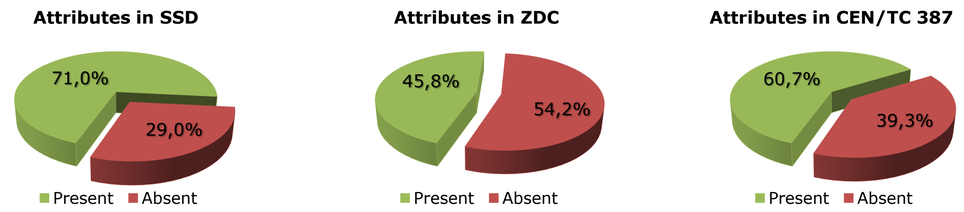


Figure 1 – Percentage of overall attributes by data model.

## Discussion and Conclusion:

The attributes corresponding to aggregated values were not used to calculate percentages in figure 1 because the values must be calculated from raw data not included in the studied models. The analysis of the remaining attributes show that:

- SSD is the most complete data model (71,0 % of overall attributes), although it does not consider the attributes related with FBOs, zoonotic agents and susceptible animal population (farm);
- ZDC is the only data model that includes attributes for FBO notification, nevertheless it lacks several relevant attributes related to sample description;
- CEN/TC 387 includes 60,7 % of the overall attributes and should be further developed essentially in order to take account of microbiological food information; this document was used as reference to harmonize all the nomenclature across the current work.

This work allowed to conclude that the creation of a unique data model for PortFIR CMFI is achievable, also it suggested that the integration by EFSA of SSD and ZDC in one document is possible and desirable.

## Significance:

This work was the first stage in the development of a unique data model for CMFI.

The biggest advantage of such a data model is the ability to store all the information needed in a single database whose compatibility with the studied relevant data models allows to exchange and to share information with national and international partners, contributing to improve food safety at a global level.

## Future Work:

Build up a unique data model for CMFI.

## References:

1. Machado, C., Santos, M., & Oliveira, L. (2010). **Compilação de Dados de Composição de Alimentos para o Portal PortFIR**. *Reunião Anual PortFIR 2010.* Lisboa.

2. European Food Safety Authority. (2010). **Standard sample description for food and feed**. *The EFSA Journal*: 8, 1457.

3. European Food Safety Authority. (2011). **Manual for reporting on Zoonoses, Zoonotic Agents and Antimicrobial Resistance in the framework of Directive 2003/99/EC and of some other pathogenic microbiological agents for information derived from the reporting year 2010**. *Supporting publication 2011*: 135.

4. European Food Safety Authority. (2011). **Manual for reporting of food-borne outbreaks in the framework of Directive 2003/99/EC from the year 2010**. *Supporting publications 2011*: 138.

5. Technical Committee CEN/TC 387 "Food Data". (17/02 /2010). **CEN/TC 387 prEN_16104 Food Data – Data structure**. *Secretariat: SIS*.