FOOD PRODUCTS DATABASE: TRANSNATIONAL APPLICATION PROBLEMS

L. M. BUGYNA, O. V. PALLAH, T. V. MELESHKO, V. V. BATI, N. V. KOVAL, N. V. BOYKO

Uzhhorod National University, Ukraine

E-mail: larina.bh@gmail.com

The purpose of this study was to analyze and compare the existing international methods applied for classifying and identifying foods characteristics in a number of modern computerized databases such as Foodex, INFOODS/FAO, EuroFIR, and present our initial results of Regional and National Ukrainian food composition databases creation.

In this study theoretical synthesis and deductive analysis were used, literary review of foreign scientific peer-reviewed sources, characteristics of the software LanguaL, DaRiS were presented.

The demand for a language-independent thesaurus (LanguaL) and the needs for a practical, field-based food system (INFOODS) led to the attempts to link these tools and create a minimal set of standards and a consistent approach for the food products identifying and analyzing around the world. The examples of this combined approach were "systems mapping" and the "International Interface Standard for Food Databases".

The exploitation of different tools for compiling of the first Regional (100 local products, project) and National food composition databases (53 products of 6 prioritised traditional foods within BaSeFood projects) were reported.

Key words: food composition data bases (FCDB), LanguaL, DaRis, INFOODS/FAO, EuroFIR.

The preparation of an international systematized and unified database of reliable food data requires precise nomenclature and detailed description of goods indicating their place of origin, energy and nutritional value, organoleptic properties and a number of other important hygienic and microbiological indicators [1]. Even the verified data structured according to the existing quality standards, can be a source of error, if they are derived from products, place in the single hierarchy of which has not been determined yet. To this date, there is a single international agreement of scientists on the importance of food nomenclature and food description. Preparation of reliable food data requires precise identification of the products definitions, methodology harmonization, analytical data validation, and comparison of software used.

Thus, the aim of this paper was to analyze and compare existing international methods of classification and identification of the main food characteristics in a number of most popular modern computerized databases: Foodex, INFOODS/FAO, EuroFIR and to present our initial results of Regional and National Ukrainian food composition databases creation according to the demands of Codex Alimentarius and the first results of national and regional applying of different tools namely LanguaL and DaRIS for the creation of correspondingly National and Regional Ukrainian FCDB in order to connect it with the international food indexation resources — INFOODS/FAO and EuroFIR.

The data of LanguaL, DaRiS application, theoretical synthesis and deductive analysis, and literature review of available research data were analyzed in our paper [2, 3].

The name of the food that is included in a particular database can be ambiguous because of the linguistic or geographical features of its origin. The definition, which is displayed in the web search engine of thesauruses can't be precise. Moreover, the same names for food with different scientific terms can be used in different regions. Some countries didn't manage to recognize certain terms that are used by people in other parts of the world or even in the same country as well. The situation is even more complicated by homonyms, synonyms, identical (consonant) trademarks for different products, as well as culinary or technological conditions of their production.

Since most databases use different methods of product identification, it is difficult to imagine an objective exchange of data between countries, between organizations within the same country, or even between employees in the same institution. That is why, this article laconically examines the existing identification systems used in the databases on food composition and characteristics, as the authors believe, that the international understanding of standardized food identification can enable to solve many problems arising from misunderstandings in this area.

Food classification systems. Earlier, there were two separate and seemingly opposite methods of solving food identification problems: products were classified by the "universal" general categories, or single descriptions of individual products were present in the databases. At present, the first approach has evolved and classifications have turned into complex hierarchies. Now we have many standardized classifications that have certain relevant legal documentation and are reflected in the thesaurus. Most of the national and regional databases use country-specific food classification systems which had been developed on the basis of national criteria, therefore, many specific food groups are observed.

This issue touches upon many legal aspects, highly depends on traditions of each country that, along with economic and cultural importance, actualizes this importance dramatically. For example, there is a separate group for coconut products in the databases developed in the Pachi Islands, and groups of different types of bananas, corn and corn bread in the database of Central America and Panama, a group of edible insects in the database of Thai food compositions.

National or regional classification systems are often difficult to use as an international basis, so we will not consider them in this article. Let us proceed to the known unified bases that can be applied to many modern cultures.

All these food classification systems have been developed for general information purposes. Their existing food codes are not specific and cannot replace national codes in the integrated food composition or value databases. All of the above systems refer to "classification" category. All of them were created for different purposes and reflect the peculiarities of the legislation of different groups of countries. To systematize them together does not appear to be real empirically. So demonstrating the enormous difference underlying these systems, one can see that, for example, when classifying cheeses into categories, the CIAA system differentiates them, first of all, as unripened, ripened, processed cheese. Eurocode-2 classifies cheeses, first of all, according to their consistency (soft, hard), and then, according to their fatness. PROCOME classifies all cheeses by "natural pure cheese" and "CCPR cheese" categories (with residues and contaminants). Classifications, even within a single system, can be contradictory, and their existence proves that there can be no single International classification system that can be unanimously approved and regulated. In other words, there is no single classification, which would be able to meet the needs of any food composition database compiler.

Next, we will consider systems of product identification in the special databases, which function according to the internal codes and descriptions.

The FAO/INFOODS Global Database for Pulses on Dry Matter Basis (PulsesDM1.0) provides nutrient values for pulses on a DRY MATTER BASIS — it is intended mainly for standard setting purposes [4]. Pulses are a subgroup of legumes that includes dry edible seeds with low fat. The data were recalculated to dry matter basis from the average nutrient values for 16 species published in the FAO/INFOODS Global Food Composition Database for Pulses (uPulses). In uPulses, data derived mainly from chemical analysis, complemented by data from other published sources and compiled following standards and guidelines outlined by FAO/INFOODS. PulsesDM cover minerals, vitamins, phytate, amino acids

and fatty acids fractions for foods in raw and processed forms. The data are presented in Excel format alongside with a comprehensive documentation in PDF format [5].

EuroFIR AISBL, an international, member-based, non-profit Association under Belgian law, was set up in 2009 to ensure sustained advocacy for food information in Europe. Its purpose is to develop, publish and exploit food composition information, and promote international cooperation and harmonization of standards to improve data quality, storage and access. EuroFIR AISBL draws together the best available food information globally from 26 compiler organizations in Europe, USA and Canada (FoodEXplorer) as well as validated information about bioactive compounds (eBASIS).

Food composition tables were originally produced as printed versions, and for many years this remained the only format. However, computerized databases have become increasingly important because they can hold large amounts of data and allow easy access to and manipulation of data. In more recent development, being facilitated and encouraged within Europe by EuroFIR, many national databases are now available online. A wide range of nutritional analysis software is also available [6].

In 2013, European compilers produced a food composition dataset for *EFSA* that aimed to provide an updated food composition database covering approximately 1750 foods and to expand the dataset to include harmonized information on the most common composite recipes of European countries. The dataset has been compiled to be compatible with the EFSA Guidance on Standard Sample Description for Food and Feed [7] and included additional descriptors from the EFSA FoodEx2 classification system [8].

Even using the most comprehensive and well documented food composition the databases do not guarantee robust and reliable results, as there are many errors that can arise in using food composition data. These include errors in matching foods, use of incompatible data, inappropriate strategies for dealing with missing values, and problems relating to the use of nutritional analysis software. Using food composition data to estimate nutrient intakes or the nutrient content of a recipe or menu can yield further errors owing both to the limitations of dietary assessment techniques and to errors associated with

dietary assessment (e.g. conversion of reported portion descriptions to weight).

The EuroFIR FoodEXplorer facility is an innovative interface, which can be accessed online and allows its users a simultaneous search of standardized and specialized food composition databases (FCDB). Users have access to a wide range of European data, foods and nutrients through harmonized data description and associated nutrient value information.

Food classification and food description are completely different categories, since they have different purposes; however, sometimes they are "mixed" into a single whole. Classification systems have only one tendency — to group products with similar characteristics (and then, not always objectively), which means, that it is a tool of the end user of the data. Description system is a data source tool that gives a description of food as accurately as possible.

The first system that carried out sequential indexing and search of food data, carried out by means of specially built in thesauruses was the INFIC/ENFIC system for animal feed identification. There, a vocabulary control is achieved by deliberately limiting the scope of terms and through its direct reflection of hierarchical relationships. The structure allows making changes by adding new "points of view" to describe food, incorporating new information. Thus, the thesaurus is well adapted to product features that can change over time.

LanguaL thesaurus, which is used in the USA, Europe and is projected on numerical food databases, is the optimal for a person "vocabulary" existing today (Fig. 1). At first, LanguaL was called as "food factorization dictionary" (McCann et al., 1988) and was created at the end of 1970 by the Center for Food Safety and Applied Nutrition, with the help of the USA Food and Drug Administration (Hendricks, 1992). Since 1996, the European Technical Committee has taken charge of the thesaurus. In total, more than 40,000 food products have been described in different countries using this dictionary. LanguaL is a multi-modal, convenient and multilingual thesaurus organized in 14 different branches, characterizing the nutritional and/or hygienic qualities of food products, which is essential for biological and medical research.

The basic concepts of LanguaL are that:

1. Any food (or food product) can be systematically described by a combination of characteristic;

How is LanguaL™ used to describe food?
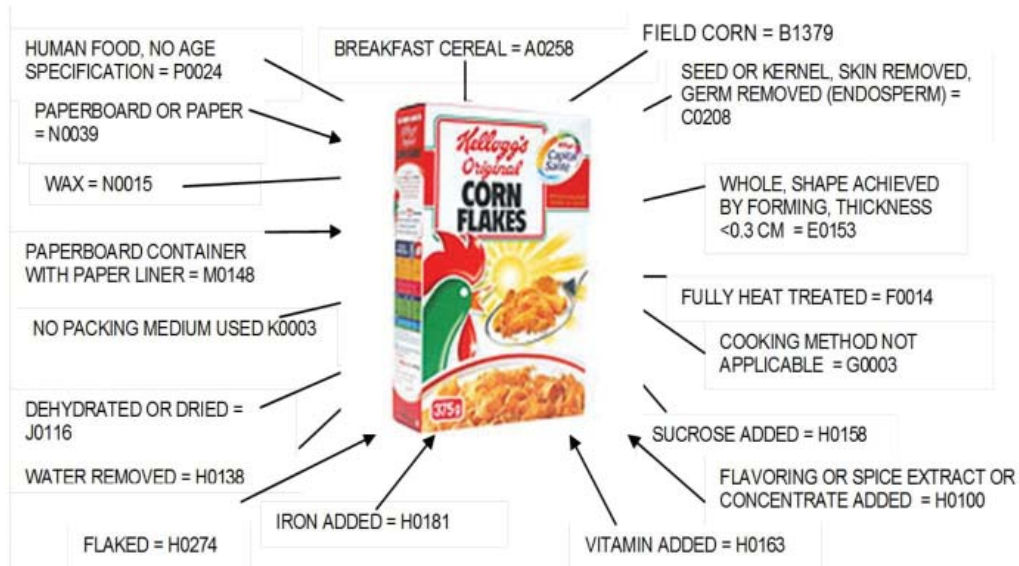
- Descriptors are chosen from each facet

HUMAN FOOD, NO AGE SPECIFICATION = P0024

PAPERBOARD OR PAPER = N0039

WAX = N0015

PAPERBOARD CONTAINER WITH PAPER LINER = M0148

NO PACKING MEDIUM USED K0003

DEHYDRATED OR DRIED = J0116

WATER REMOVED = H0138

FLAKED = H0274

IRON ADDED = H0181

BREAKFAST CEREAL = A0258

FIELD CORN = B1379

SEED OR KERNEL, SKIN REMOVED, GERM REMOVED (ENDOSPERM) = C0208

WHOLE, SHAPE ACHIEVED BY FORMING, THICKNESS <0.3 CM = E0153

FULLY HEAT TREATED = F0014

COOKING METHOD NOT APPLICABLE = G0003

SUCROSE ADDED = H0158

FLAVORING OR SPICE EXTRACT OR CONCENTRATE ADDED = H0100

VITAMIN ADDED = H0163

Fig. 1. **Description of food by LanguaL**

2. These characteristics can be categorized into viewpoints and coded for computer processing;

3. The resulting viewpoints/characteristic codes can be used to retrieve data about the food from external databases [9].

Each descriptor has a unique base code that points to equivalent terms in different languages, which makes the thesaurus independent of language features and peculiarities. For the past 2 years, the thesaurus has been significantly modified and now it provides open links to international food categories and coding systems. The official international version of the thesaurus was published on the LanguaL website (http://food.ethz.ch/langual), where copies of the thesaurus are available on request. The interface allows searching for products in American, Danish, French, Hungarian and a number of other databases, which maximally facilitate the exchange of information at the transnational level. However, some aspects still require further clarification, as LanguaL lacks some of the food groups that are used in the national tables. There is also a need to optimize software for searching and indexing relevant terms. The European LanguaL Technical Committee is currently working on these issues (Fig. 2).

By food description, the *INFOODS* management board prepared cognominal INFOODS system with the support of the Committee on Food Nomenclature and Terminology in 1987. The purpose of the INFOODS nomenclature system is to provide a basis for data exchange between primary sources and compilers of systems devoted to information on food composition. The system is a wide, multifaceted and open mechanism. The INFOODS management board offers criteria for determining whether the food is one-component or multi-component and provides different sets of descriptive aspects for these two large classes. However, this thesaurus is significantly inferior to LanguaL in its completeness. Also, it does not provide an indexer/retriever with a list of possible terms (synonyms) for any product of interest. A draft of the increased number of terms was repeatedly prepared, but it has not been published yet. The INFOODS system or its individual forms is used in New Zealand, the South Pacific, several African countries and 10 Latin American countries. LanguaL is common in other countries.
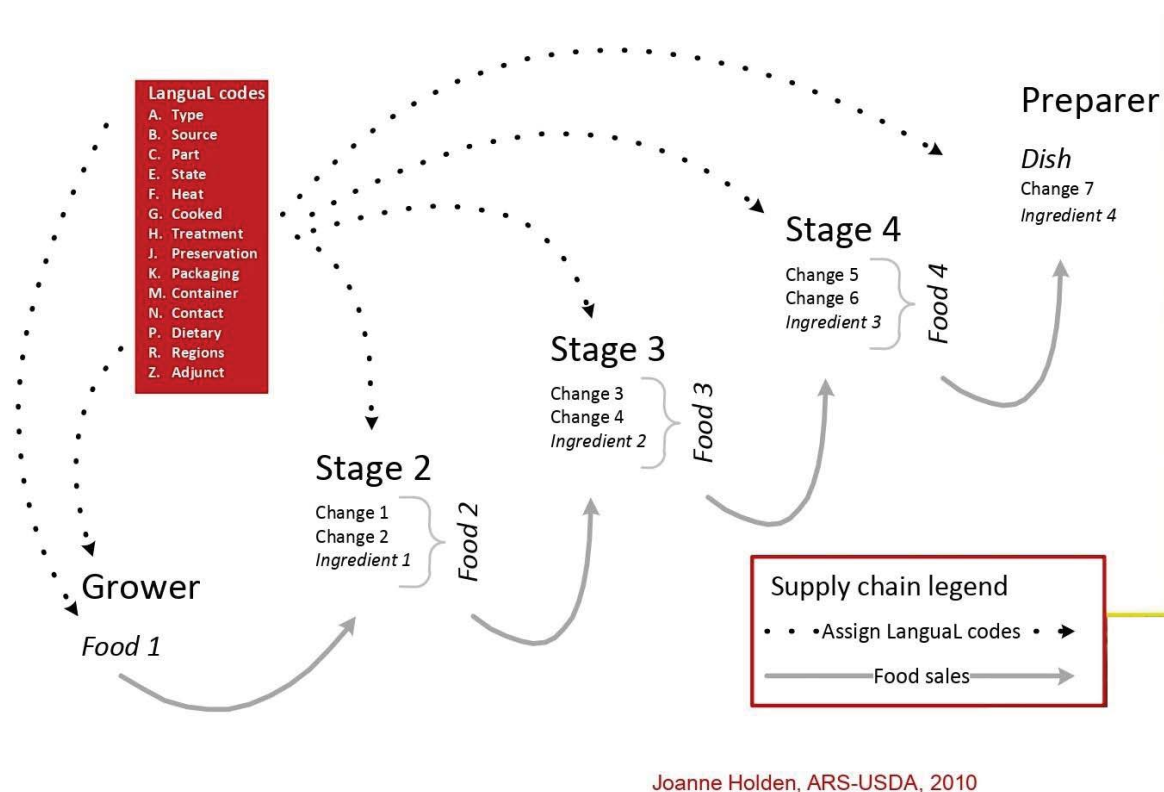
LanguaL codes
A. Type
B. Source
C. Part
E. State
F. Heat
G. Cooked
H. Treatment
J. Preservation
K. Packaging
M. Container
N. Contact
P. Dietary
R. Regions
Z. Adjunct

Preparer

Dish
Change 7
Ingredient 4

Stage 4

Change 5
Change 6
Ingredient 3

Food 4

Stage 3

Change 3
Change 4
Ingredient 2

Food 3

Stage 2

Change 1
Change 2
Ingredient 1

Food 2

Grower

Food 1

Supply chain legend
• • • Assign LanguaL codes • • ➤
Food sales ➤

Joanne Holden, ARS-USDA, 2010

Fig. 2. **LanguaL and the food supply chain (Joanne Holden, ARS-USDA, 2010)**

Earlier, several other test food thesauruses were created to manage bibliographic information: CAB thesaurus, which is used by Nutrition Abstracts and Reviews; IFIS thesaurus, which is used in the theses of Food science and technology society; AGROVOC thesaurus (FAO, 1998), Created on FAO AGRIS and CARIS data banks. However, they were developed primarily for documentation purposes and do not have the specificity of product descriptions, which only LanguaL has.

Thus, there are two main systems that are used to describe food in food composition and characteristics databases: LanguaL thesaurus with international terms and the INFOODS system, which application depends on the national language. A comparison of these two systems was carried out by the coordinators of the regional data centers (Burlingame, 1998). LanguaL thesaurus language scored higher in relation to solving problems of the language barrier and culture, which is the reason for its unconditional adoption in Europe. However, the INFOODS system has shown some better results regarding the relationship of data compilers and local utility for ordinary users, who want to obtain data on the composition of consumable food quickly and easily. The INFOODS system is easier and faster to use and does not require searching for complex terms and codes in lists. The demand for a thesaurus that is language-independent (LanguaL), and the need for a practical, "field" system (such as the INFOODS system) for food products, have led to attempts to combine these systems and create a minimum set of standards and an agreed approach for food identification around the world. Examples of this combined approach are "System mapping" and "International Interface Standard for Food Databases".

For the first time the Ukrainian National and Regional Food Composition Databases were created by us. In the food industry, access to food database opens up new opportunities for selection of food components for the analysis of available data, creating new recipes that will have a positive effect on production of foods. Also, it is very important for Ukrainian consumers to have access to the information about the quality and composition of foods. Especially important is the content of allergens, sugar, GMOs, information about shelf life of products, etc. To ensure consumer access

9

to such information scientists have to make a huge effort on creation of database of food in Ukraine [10]. One of the major problems in Ukraine is lack of reliable verified information on the composition of food products due to non-compliance by manufacturers with the relevant labeling rules. Also analytical stage where the main components of food products are analyzing have a big value in food quality control. Nowadays we do not have an unified analytical system, methods of food components analyzing and also qualified employees and compilers.

Thus, the first breakthrough in the international food identification became apparent after the scientific recognition of the benefits of using a multidimensional approach to food systematization in food databases. The second breakthrough is the recognition of the need for an alternative classification/descriptive system that would combine the advantages and exclude the disadvantages of the program products described above.

Work in the field of transnational food identification in databases is carried out with the help of the IUNS/FAO international target group and was approved at the third International Conference on Food Data. This target group will continue to review and analyze the work that will be carried out on existing food classifications and descriptions in order to harmonize the international use of the final result.

Nowadays, the Global Harmonization Initiative (GHI) is created in Europe. GHI is the international non-profit network of individual scientists and scientific organizations working together to promote harmonization of global food safety regulations and legislation and this is not possible without reliable information on the composition of food and its availability.

REFERENCES

1. *Almli V. L., Verbeke W., Vanhonacker F., Næs T., Hersleth M*. General image and attribute perceptions of traditional food in six European countries. *Food Quality and Preference*. 2011, 22(1), 129–138. https://doi.org/10.1016/j.foodqual.2010.08.008

2. *Porubská J., Giertlová A., Morochovičová M., Kováčiková E., Porubský O*. The Slovak national food composition database: New management system DaRiS. *Journal of Food Composition and Analysis*. 2014, 34(1), 26–38. https://doi.org/10.1016/j.jfca.2014.02.007

3. *Kapsokefalou M., Roe M., Turrini A., Costa H.S., Martinez-Victoria E., Marletta L., Finglas P*. Food Composition at Present: New Challenges. *Nutrients*. 2019, 11(8), 1714. https://doi.org/10.3390/nu11081714.

4. *De Victoria Muñoz, E. M*. Do we really know the composition of our foods? *Nutricion hospitalaria*. 2013, 28(5), 73–80.

5. International Network of Food Data Systems (INFOODS). Available online: http://www.fao.org/infoods/ infoods/en/ (accessed on 20 May 2019)].

6. EuroFIR. Network of Excellence and NEXUS Projects. Available online: http://www.eurofir.org/our-resources/ noe-and-nexus-projects/ (accessed on 20 May 2019). https://www.mdpi.com/2072-6643/11/8/1714/html

7. European Food Safety Authority. Standard sample description for food and feed. *EFSA J*. 2010, 8, 1457.

8. European Food Safety Authority: Food Classification. The Food Classification and Description System FoodEx2 (Revision 2). 2015. Available online: http://www.efsa.europa.eu/en/datex/datexfoodclass.htm (accessed on 20 May 2019).

9. LanguaL™ - the International Framework for Food Description Available online: https://www.langual.org/default.asp

10. *Koval A., Pauk L., Buhyna H., Costa T., Albuquerque A, Sanches-Silva J*. Development of national food composition database in Ukraine following EuroFIR and BaSeFood projects Plumb, M. Roe, P. Finglas, M. Glibetic, N. Boyko. *Traditional food international*, 4–5 October 2012, P. 89.

БАЗИ ДАНИХ ХАРЧОВИХ ПРОДУКТІВ: ПРОБЛЕМИ ТРАНСНАЦІОНАЛЬНОГО ЗАСТОСУВАННЯ

Л. М. Бугіна, О. В. Паллаг,
Т. В. Мелешко, В. В. Баті,
Н. В. Коваль, Н. В Бойко

Ужгородський національний університет,
Україна

E-mail: larina.bh@gmail.com

Метою цього дослідження було проаналізувати і порівняти існуючі міжнародні методології, що їх застосовують для класифікації та ідентифікації характеристик харчових продуктів у сучасних комп'ютеризованих базах даних (Foodex, INFOOD/FAO, EuroFIR), а також подати наші первинні результати створення регіональних і національних українських баз даних за складом продуктів харчування.

У дослідженні використовували теоретичний синтез і дедуктивний аналіз, огляд літератури зарубіжних наукових рецензованих джерел, LanguaL, DaRiS.

Попит на незалежний від мови тезаурус (LanguaL) і потреба в практичній, польовій системі харчування (INFOODS) спонукали до спроб зв'язати цю систему та створити мінімальний набір стандартів і послідовний підхід до визначення харчових продуктів у всьому світі. Прикладами цього комбінованого підходу є «Картографування системи» та «Міжнародний стандарт інтерфейсу для баз даних стосовно продуктів харчування».

Уперше впроваджено різні інструменти для складання перших регіональних (100 місцевих продуктів) та національних баз даних про склад продуктів харчування (53 продукти із 6 пріоритетних груп традиційних страв і напоїв) у проекті BaSeFood.

Ключові слова: композиційні бази даних харчових продуктів (КБДХП), LanguaL, DaRIS, INFOODS/FAO, EuroFIR.