Food Chemistry 238 (2018) 166–172

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

FoodCASE: A system to manage food composition, consumption and TDS data

Karl Presser *, David Weber, Moira Norrie

Department of Computer Science, ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Article history:
Received 30 June 2016
Accepted 18 September 2016
Available online 19 September 2016

Keywords:
Data management system
Food composition data
Food consumption data
Total Diet Study (TDS) data

ABSTRACT

Food and nutrition scientists, nowadays, need to manage an increasing amount of data regarding food composition, food consumption and Total Diet Studies (TDS). The corresponding datasets can contain information about several thousand different foods, in different versions from different studies. FoodCASE is a system that has been developed to manage these different datasets. It also support flexible means of linking between datasets and generally provide support for the different processes involved in the acquisition, management and processing of data. In this paper, the most important concepts to implement existing guidelines and standards for proper food data management are presented, as well as different use cases of data import and proofs of concepts demonstrating the ability to manage data in FoodCASE.

1. Introduction

Switzerland has a long history of developing computer systems for food composition data. The first system was developed in 1992 with a DOS user interface, which was later updated to Microsoft Access, while the second system was planned to be a virtuoso in providing different interfaces. One of the planned interfaces was with the landline telephone, in which the dial pad could be used to navigate a system narrating food composition information. Two other interfaces were planned with fax machines and email, in which the user could send requests to the system, and the system could respond with an answer. However, the only plans implemented were a website and interfaces.

Two additional systems were created before the development of FoodCASE started in cooperation with Swiss compilers, the Swiss government and the European Food Information Resource (EuroFIR) project (EuroFIR, 2013). It was decided to create FoodCASE because the technology used in the former systems was considered outdated, and because the new European Committee for Standardization (CEN) standards and recommendations from the EuroFIR were meanwhile established (Becker, Unwin, Ireland, & Møller, 2007) (Becker et al., 2008). Further developments of FoodCASE where later implemented in close cooperation with food compilers across Europe.

In 2014/15, Switzerland conducted the first national food consumption survey, and FoodCASE was chosen as the storage system for the interview data. The main goal was to keep the consumption and food composition data in the same system, to enable automated linkage between the two datasets. Portugal also wanted their food consumption information kept in FoodCASE, which requested the implementation to be more general so that different food consumption datasets can be managed.

The TDS-Exposure project (TDS-Exposure, 2016) aimed to harmonise the analysis of dietary contaminants throughout Europe, and to create an EU-wide network of Total Diet Study (TDS) centers. FoodCASE was further extended in this project to manage also TDS data. TDS data is similar to food composition data in terms of value documentation of analysed values. However, it is different in that foods collected for TDS are pooled before they are analysed, and more sampling information needs to be stored. The main difference in terms of data management is that for food composition the focus lies on the storage of analytical values and generation of new values whereas in TDS, the focus is on value documentation for exposure assessment.

In Section 2, we begin with a review of the existing standards and recommendations in the different food data areas. Section 3 describes the general architecture of FoodCASE, including the different datasets for food composition, consumption and TDS data. In Sections 4–6, the functionalities of the different dataset modules will be described, and feasibility studies for importing and managing the data will be presented, while the concluding remarks will be given in Section 7.

* Corresponding author at: ETH Zurich, Department of Computer Science, CNB E 102.2, Universitätsstr. 6, 8092 Zurich, Switzerland.

E-mail address: karl.presser@inf.ethz.ch (K. Presser).

http://dx.doi.org/10.1016/j.foodchem.2016.09.124
0308-8146/

2. Background

There are many European and international initiatives to food composition data, which must be considered when implementing a food composition database system. These contributions cover the data levels in which the food composition data is processed, the entities and attributes, food descriptions, component identification, documentation of nutrient values and data interchange formats. The following review is divided into these topics.

2.1. Data level for food composition data

According to Greenfield and Southgate food composition data exists at four different levels (Greenfield & Southgate, 2003): data source, archival data, reference database and user database.

The data source level includes published and unpublished research papers and laboratory reports that contain analytical data. If the data from these multiple sources are brought together in a single collection, then the data is said to be at an archival data level. The archival data should be held in the units in which it was originally published, and any relevant and available information should be extracted from the original publication and stored with the archival data so that referring back to the data source becomes unnecessary. The data level of a reference database is usually part of a computer database management system, which should be a complete pool of the scrutinised data, but with the data converted into uniformly expressed standard units and nutrients expressed uniformly. The nutrient values from the archival data are brought together for each food, and possibly, enriched by adding information such as density, pH value or non-editable portion. This level also contains the calculated, combined, averaged and weighted values, with storage of the conversion factors, calculations and recipes. The fourth level user database is generally a subset of the reference database, and contains selected or combined data. The defining feature of a user database may be that it provides one series of data per food item.

EUROFOODS took over this four data level model in their recommendations for food composition database management and data interchange (Schlotke et al., 2000a), which was based on the findings from the European Cooperation in Science and Technology (COST) Action 99 project (Schlotke et al., 2000b). At the data source level, an addition was made suggesting that the data might be managed within a laboratory information management system (LIMS).

In general, this four data level model is a good proposal, but there are some issues to consider when implementing a food composition database. For example, the data source level can be out of scope if the compiling Organization does not have own laboratories. However, even if it has a designated laboratory, some of the data does come from other sources; therefore, the data source level is not normally in the compilation software. The description of a reference database (level three) suggests that data for individual analyses should be held separately, and if a system is able to calculate, query, edit, combine, average and weigh the values, these calculations should also be stored. On the other hand, the user database (level four), in some cases, can contain weighted or averaged values to ensure that the values are representative of the foods in terms of the use intended. These values may be amalgamated, rather than shown as individual constituents.

The issue here is that the distinction between levels three and four gets a bit fuzzy. Amalgamated values can already be calculated at level three, or first at level four. Another issue is that level three could contain results of individual analyses in standardised units, as well as averages or weighted averages. This issue is of a more practical nature, because a mix of individual analyses and average values can make it harder to find the correct value that should be propagated to level four, or chosen for a recipe calculation. There is also a redundancy issue in levels two, three and four; for example, all of the values from level two also exist in level three as converted values in standardised units. If the averages are calculated and stored in level three and propagated to level four, there is additional redundancy. Moreover, an issue was found during our work with the Swiss compilers and some of the European compilers. Most of the compilers do not have the budget or time to manage their data over three levels, with most of them storing their data in one level without value documentation, if only for pragmatic reasons. Therefore, a food composition database system should offer an efficient and feasible solution for data management while accommodating the different needs.

2.2. Data schema for interchange

The EUROFOODS/COST 99 recommendations include the complete food composition data schema for a relational database. The recommendation also contains a predefined set of attributes for each entity. This rather complex and circumstantial data schema was simplified in the EuroFIR project into the EuroFIR standard (Becker et al., 2007) (Becker et al., 2008). A CEN standard as well as an XML interchange schema, called the Food Data Transport Package (FDTP), were then generated based on the EuroFIR standard (Becker, 2010) (CEN, 2012) (Møller & Christensen, 2008).

The EuroFIR standards suggest that the schema implementation is a design matter for the individual national databases, whereas the main goal is to define the data interchange format. Nevertheless, the data schema does not differentiate the data levels from the previous section, and the level specific attributes are mixed in the entities. In addition, the entity value contains certain of the method's attributes, although the method specification is its own entity. Moreover, the EuroFIR standards have not been updated recently, so the new European Food Safety Authority (EFSA) food classification systems FoodEx and FoodEx2 are missing. On the other hand, each facet of the food description system LanguaL has been incorporated into the food entity while assignments of the LanguaL codes could be enough.

2.3. Food classification and description

A classification system normally has distinctive categories describing the main characteristics, whereas a food description system describes food items based on several distinctive aspects. A food description system can be looked at as a multi-faceted classification system. Several classification and description systems exist and target specific purposes, such as food composition, food consumption and exposure assessment, and comparisons of different international systems have been published (Truswell et al., 1991) (Ireland & Møller, 2000) (EFSA, 2011b). A food composition database management system in Europe must provide the EuroFIR food classification (Ireland & Møller, 2006) and LanguaL, focusing on food composition, and FoodEx2 (EFSA, 2015), focusing on exposure assessment, as just a few of the key systems. Compilers normally use their own national food categories, which is a proprietary classification.

LanguaL has several classification systems included under Facet A, and the basic idea is to integrate them. The challenges include keeping LanguaL up to date and integrating FoodEx2, which allows for the addition of country specific codes.

2.4. Other thesauri and data quality

Several systems exist in the area of component identification. The EuroFIR component thesaurus, also defined in the EuroFIR