

Challenges in Food Composition

Subjects: Food Science & Technology

Contributor: EMILIO MARTINEZ DE VICTORIA MUÑOZ

Food composition data is important for stakeholders and users active in the areas of food, nutrition and health. New challenges related to the quality of food composition data reflect the dynamic changes in these areas while the emerging technologies create new opportunities.

Keywords: food composition data ; food description ; food matching ; data quality ; dietary assessment ; research infrastructure

1. Introduction

Food composition data is important for a wide range of stakeholders and users including researchers, public food and health policymakers, healthcare professionals, industry (food, agriculture, software developers), consumers and for educational purposes. Each of these groups has different uses for the data and individual users within the groups may have very different requirements and expectations of the data. For many uses, food composition data is linked to food purchase or consumption data to enable calculation of nutrient intake and assessment of quality of the diet. Traditionally, food composition data has mainly been used for research and to underpin national dietary monitoring programs related to policy.

The major source of food composition data for both non-commercial (research, academic, public operators, educators) and commercial use is authoritative national food composition datasets. These datasets are typically produced and published by national governmental bodies but may also be produced by research institutes and other non-governmental agencies. The data may be published as either reference datasets, which may be incomplete in terms of foods and nutrients included, or user datasets that have been modified from the original reference datasets. Researchers in dietary science (and some industry users) tend to use data at a detailed level and often want to use modified or extended versions of published datasets. Research users often access composition data through datasets that have been embedded in a wide range of software tools and in many cases, users may not be aware of the original data sources or of limitations in the data provided, including whether or not the software includes up-to-date values ^[1].

Nutritional information is also available for many branded food products, particularly those produced by large manufacturers and sold by major retailers. While this data is easily available for individual products (through retailer and manufacturer websites), compiled datasets have not been widely used for research purposes. Datasets for branded products are commercially available and some manufacturers or retailers may provide product datasets, but they are either expensive and/or technically difficult to use. However, there is potentially significant research value in these datasets being more readily available for academic research purposes. The United States Department of Agriculture (USDA) has integrated a branded food database in the national nutritional database ^[2]. In addition to nutrient data, researchers are increasingly interested in using data on non-nutrient bioactive compounds. Many compounds have been found to have beneficial or toxicological health effects in humans and data on bioactive compounds has been compiled and published in online databases ^{[3][4][5][6]}. The research use of this data is rapidly increasing in response to public awareness and the wider availability of published data.

Nationally representative datasets are produced in most developed countries and increasingly in developing countries, although coverage of foods and nutrients may be more limited. Datasets are typically produced, managed and published by groups with a high level of sustainable expertise in food composition. Sources of data used within the datasets include: analytical data produced specifically for the dataset using standard methods in accredited laboratories, data from scientific or grey literature, calculations from ingredients of composite foods, manufacturers' data and data from other national datasets. Data is managed and published using a range of data handling tools, most commonly relational databases, although some datasets are still compiled and published in spreadsheet format. Although data standards do exist, the structure of food composition data is not yet fully standardized and depends on the source of the data and the knowledge and expertise of data compilers.

There have always been challenges in harmonizing approaches to acquisition and publication of nutrition, health and lifestyle data, and while data processing and publication in electronic form makes large datasets easier to handle and more accessible, the challenges related to data quality, data exchange formats and documentation have increased. Many international projects and research networks have initiated standardization of methods to collect, manage and publish food composition data, but data standardization has not kept pace with fast moving developments in information and communications technology. The ability to generate and exchange large amounts of data has therefore highlighted existing limitations in data structure. A European Food Safety Authority (EFSA) project to produce a European food composition dataset for use with the EFSA Comprehensive European Food Consumption Database highlighted limitations in combining European food composition data from different countries [7].

2. Challenges in Food Composition

Although food composition datasets and other large food and nutrition related datasets (e.g., food consumption data, food trade data) are now more accessible, and technology enables transfer of data between users, methods used to generate and publish data can limit effective use. If data are made available according to the FAIR Data Principles of findability, accessibility, interoperability, and reusability [8], many users could acquire and use data without being aware of limitations and there is a risk that these limitations may influence the way that data analysis is interpreted. In this issue, guidance by the European Open Science Cloud has been delivered and specific research groups have been recognized by the European Commission [9]. There are existing networks of expertise that can help to mitigate these limitations and they can play an important role in providing better quality data that is more likely to be fit for purpose. Food composition data can be derived from a range of sources and most food composition datasets contain values produced in a variety of different ways. Data quality is associated with various factors including: food description, component identification, sample collection, sample handling, analytical method, value documentation and laboratory performance. Systems for evaluating quality of nutrient data have been developed [10][11] but information on data quality is not usually published. Quality of food composition data has been considered as part of a Quality Management Framework for the production and management of food composition data [11].

One of the major problems of food composition datasets has always been providing and maintaining data that reflects the range of foods and nutrient composition of foods that are being consumed. The nutrient content of food changes over time and food composition databases should be continually revised to provide data for new foods and for foods where composition has changed. Nutrient composition may change for reasons including growing conditions, changing agricultural practices, plant breeding developments, changes in processed methods and changes in consumer expectations and preparation [12]. Processed foods are constantly changing as manufacturers try to protect or increase market share and profits and respond to policy changes dictated by a combination of government policies and consumer pressure, e.g., reduction of sugar, salt, saturated and trans fats. In recent years the rate of changes in composition and foods consumed has increased in many countries, in line with increased focus on the role of diet in population health. The need to maintain nutrient datasets is constant but in practice most national datasets are not continually revised and updated versions are published according to resource availability and based on foods that are identified as being the most important contributors to diets. Priorities may be based on food type, e.g., surveys of eggs, meat, fish, fruit or vegetables, or in some cases based on nutrients that are expected to have changed, e.g., sugars, fatty acids, sodium. The recent economic climate has resulted in most compilers having to work with increasingly limited resources, including budgets for analysis, equipment and staff.

The approach to maintenance of national datasets means that a percentage of data will always be old, and some old data will not be representative of foods that are currently consumed. Ideally, data contained within datasets will include information on the date of generation and/or publication and will be validated at the time of publication. Even where information on data validity is available, datasets will almost certainly still contain out of date values and, following publication, validity of the dataset will decrease over time. The impact of out of date values depends on what the data is used for but can be very significant in some cases. One of the main reasons for national food composition tables' (FCTs) data obsolescence is the lack of funded programs to a) perform analyses and b) to acquire innovative tools for making quicker chemical analysis processes [13]. This issue also causes a non-secondary effect as FCT users try to complete missing data borrowing values from other datasets, so obsolescence can be propagated. Specific strengths, weaknesses and challenges may be identified when considering national or regional FCTs and FCDBs.

The Mediterranean countries are an interesting case study because of the coexistence of European, African and Middle East countries along the Mediterranean region, which provides a good example when considering regional FCTs and FCDBs. **Figure 1** is a description of the current strengths, weaknesses and challenges of food composition datasets in the

Mediterranean region and was developed during the NUTRIMAD 2018 conference by compilers responsible for national datasets in the region.

Strengths	Weakness	Challenges
• Evolution of FCD in some countries in recent years • Participation of some Mediterranean countries in the EuroFIR consortium, providing access to resources, training and collaborative partnerships • High scientific competence of trained compilers that allows the critical evaluation of resources and data • A well studied regional model of food consumption and dietary patterns that have been associated with positive effects on health • Common characteristics in commonly consumed regional foods and dishes	• Limited resources to update data to reflect composition of currently consumed foods • Lack of standardization in building food composition databases • Difficulties coordinating with multiple bodies that may provide data, funding and services (e.g. academia, government, industry, laboratories) • Capturing cultural specificities and needs for defining a common language to describe and compare food consumption, food purchase, food preparation, recipes, composite dishes	• Establishing and maintaining a network of Mediterranean countries, to create new capacities • Obtaining funding for networking, capacity building, and generation of new data • Developing and maintaining Food Composition Database Management software for regional data • Capturing cultural differences as well as commonalities between Mediterranean Countries when building FCDB • Collecting information on the main Mediterranean recipes for calculating new composition data • Producing data to help protect authenticity and traceability of local products

Figure 1. Strengths, weaknesses and challenges of food composition databases (FCDBs) in countries in the Mediterranean region.

Shared and exploitable unique characteristics in the local food system include respect for diversity in the primary production, preservation of traditional foods and a large number of products that have been assigned as Designation of Origin (PDO) or Protected Geographical Indication (PGI) ^[14]. Synergies among the Mediterranean countries in development, update and maintenance of FCTs and FCDBs is required in order to overcome, at least partially, the limitations in resources and respond to new challenges of the digitally interconnected world. Intellectual support, sharing of data and best practices may be drawn from participation to networks such as European Food Information Resource (EuroFIR) ^[15] and International Network of Food Data System (INFOODS) ^[16].

Following the recognition of these common characteristics a consortium of six Mediterranean Sea coast countries (Cyprus, Croatia, Spain, Greece, Italy, Morocco) and Portugal succeeded in 2013 to include the Mediterranean Diet in the list of Intangible Cultural Heritage of UNESCO ^[17]. This initiative emphasizes and shares internationally the uniqueness, values, and collaboration of Mediterranean countries and creates the potential for exploitation of this dietary model to the benefit of local economies and culture.

Published food composition datasets can be either reference or user datasets ^[11]. Published datasets are usually a sub-set of reference datasets and may not include all foods and nutrients included in the reference dataset. Some published values may be recalculated for publication of derived values, including recipe calculations and aggregations of branded food data. Reference datasets and user datasets often do not include values for all nutrients for all foods so values may be added to fill gaps. These values may be imputed from other values, estimations from other foods or data sources, or may be logical zeros (e.g., cholesterol, vitamin B12 in plant foods). Some nutrient datasets and bioactive databases only include data for raw foods and do not provide data for cooked and processed foods. These datasets are unsuitable for use in intake calculations because they do not include data for many foods as consumed and do not include data for composite dishes (foods prepared from more than one ingredient). For most purposes it is necessary to adapt published data for specific uses. Most uses will require values for all nutrients to avoid the possibility of missing values being assumed to be insignificant and treated as zero. Some nutrients, e.g., fatty acids and individual sugars, are usually only provided for a limited sub-set of foods so they may also need to be added by users. National datasets do not all include the same nutrients and while the major nutrients will be included, nutrients such as individual sugars, individual fatty acids, and some minerals, trace elements and vitamins may not be included. This is not a problem when users are only using data from one country but can be problematic when data from different countries and sources are combined. Many users of the data are not aware of the limitations of published data, despite detailed documentation being provided for most national datasets. Where users are aware of the need to adapt data, there is often a gap in the knowledge and skills and/or the resources needed to adapt the datasets for correct use.

Datasets of branded foods may be updated continually following agreements between producers and wholesalers and retailers, who often require updated composition data as a pre-requisite for distributing or stocking products. This information is often supplied by a third-party organization that receives information from producers and compiles the data into datasets that are then supplied to retailers for online use. An example is Brandbank ^[18] who compile and distribute data globally, on a commercial basis, to provide online solutions for retailers. Although the values included in datasets are as up-to-date as possible, the format of the data and consistency of food description pose some challenges to users. The data relate to very specific products, based on Global Trade Item Number (GTIN) code ^[19], and new products are

continually added while others are discontinued. Branded food data usually only provides nutrient values that are required or permissible for food labelling. The EU labelling regulations ^[20] do provide a standard for how the values are calculated and presented but provision of values for some nutrients (e.g., fibre, starch, mono and polyunsaturated fatty acids) is optional so values for those nutrients will not be provided for all foods. Values for minerals and vitamins can only be included on food labels when a significant amount of the nutrient (15% of nutrient reference value for foods or 7.5% for beverages) is present and, as a result, mineral and vitamin contents of foods are rarely available for branded foods. For uses that require values for those nutrients, e.g., calculation of nutrient intakes, these values must be added to the dataset. It is possible for branded foods to be mapped to generic foods so that mineral and vitamin contents can be estimated but that work requires considerable expertise and technical data management skills. For that reason, branded datasets have not been completed (in terms of nutrient coverage) and fully merged with generic national datasets.

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 was compiled to be compatible with the EFSA Guidance on Standard Sample Description for Food and Feed ^[21] and included additional descriptors from the EFSA FoodEx2 classification system ^[22]. To ensure a complete nutrient dataset was provided, where data was not available in a national database, values were borrowed from another country. The project highlighted a number of significant limitations in the dataset that were related to the need to standardize approaches to compiling and using data to ensure that data is as comparable as possible ^[7].

Bioactive compound databases are compiled in a similar fashion to nutrient databases, employing a systematic search and selection process utilizing appropriate online searching tools, such as Web of Science. Published databases cannot be described as comprehensive, primarily because new literature is constantly being published and analytical methods for many bioactive components are constantly developing and are not standardized. The main sources for polyphenol composition data (USDA ^[3], eBASIS ^[4], Phenol-Explorer ^[5]) are expert-based curated databases rather than being compiled in an automated way. A dataset of anthocyanins in foods consumed in Australia has also been developed ^[6]. The literature based, curated approach means that these datasets are labor-intensive to produce, maintain and update, which in turn means they require consistent funding and resources and therefore updating may be limited.

The limitations of compiling datasets mean that there may also be limitations for users that depend on the intended use. 'Static' datasets that are included in applications may be out of date when they are used and in many cases data users do not routinely update to newer versions of the data. While updating is technically possible, many users do not have the skilled IT resources or sufficient knowledge of the data to make the necessary updates. For some purposes, e.g., research studies over extended time periods, it may not be desirable to use updated data because nutrient intake may then be changed because of changes in the underlying composition dataset rather than just in the foods consumed.

To overcome these limitations and challenges it is essential that datasets are documented as fully as possible, using international standards, to allow the possibility of reliably exchanging information between datasets. While it is increasingly easy to store and transfer large datasets and make them available online, it is usually necessary to do a lot of work with cleaning and standardizing datasets before data can be accurately and usefully combined. This work relies on the knowledge and training of a relatively limited number of compilers who have the required experience.

3. EuroFIR

Traditionally there has not been a standard structure for food composition data because datasets have been compiled independently for publication in country specific printed tables in books and scientific journals. Since the introduction of computerized data compilation and publication, there has been a trend towards more standardized data structures and control of data quality through clear documentation of the data.

There have been many collaborative projects and networks of food composition data compilers that have aimed to improve consistency and harmonization of composition databases, so that values from different datasets are of comparable quality. European projects such as EuroFOODS, Cost Action 99, the IARC European Nutrient Data Bank project ^{[23][24]} and the work of INFOODS ^[16], all made progress towards more standardized production, compilation and management of data. These and other related projects were used as the basis for the European Food Information Resource (EuroFIR) project.

The EU FP6 and FP7 EuroFIR Network of Excellence (NoE) (2005–2010) and EuroFIR NEXUS (2011–2013) projects ^[25] aimed to standardize and harmonize food composition data in Europe through improved data quality, database

searchability and standards. To further standardize the EuroFIR quality approach, new or existing procedures and tools were developed or adopted for data interchange, food description, component identification, value documentation, recipe calculation and quality evaluation of values. Following the NEXUS project, EuroFIR AISBL was formed in 2009 as an international, member-based, non-profit Association and aims to ensure sustained advocacy for food information in Europe and facilitate improved data quality, storage and access, and encourage wider applications and exploitation of food composition data for both research and commercial purposes.

4. Collaborative Networks

Collaborative networks of food composition data compilers and users have proved to be an effective approach in dealing with the challenges of producing high quality data and creating and maintaining tools to produce and manage data. Most of these European initiatives have been enabled by EU funded projects with other sources of funding allowing developments to be extended to other areas, e.g., World Health Organization, UK Global Challenges Research Funding program [26]. In addition to funding that enables data to be produced, developed and improved, it is crucial that networks are maintained to enable co-operation, training and exchange of ideas between compilers and data users from different countries and regions of the world. INFOODS and EuroFIR have been providing these opportunities and these networks continue to promote and advocate for the importance of high-quality food composition data that is relevant to specific countries and populations of consumers. EuroFIR AISBL is sustained as a not-for profit members organization that relies on a combination of funding from membership, participation in research projects and commercial activities. INFOODS is funded through the Food and Agriculture Organization of the United Nations.

Current challenges can be addressed by the formation of new networks that create synergies enabling the development and/or the continuous update of FCTs and FCDBs. For example, a Mediterranean Network expanded to include EU and non-EU countries in the Mediterranean Region would enable the linking and/or exchange of data, resources, infrastructures and best practices. The collaboration of such an initiative with EuroFIR AISBL and other networks will amplify the expected benefits thus supporting its future success and sustainability. A Mediterranean Network on FCTs and FCDBs is vital for the development of ambitious common research and policy initiatives in support of the Mediterranean Diet.

References

1. Clancy, A.K.; Woods, K.; McMahon, A.; Probst, Y. Food Composition Database Format and Structure: A User Focused Approach. *PLoS ONE* 2015, 10, e0142137.
2. United States Department of Agriculture. FoodData Central. Available online: <https://fdc.nal.usda.gov/> (accessed on 20 May 2019).
3. Haytowitz, D.B.; Wu, X.; Bhagwat, S. USDA Database for the Flavonoid Content of Selected Foods Release 3.3. Available online: <https://www.ars.usda.gov/ARUserFiles/80400525/Data/Flav/Flav3.3.pdf> (accessed on 18 February 2019).
4. EuroFIR AISBL. eBASIS: BioActive Substances in Food Information System. Available online: <http://ebasis.eurofir.org/Default.asp> (accessed on 18 February 2019).
5. Phenol-Explorer. Available online: <http://phenol-explorer.eu/> (accessed on 18 February 2019).
6. Igwe, E.; Neale, E.P.; Charlton, K.; Morton, K.; Probst, Y. First stage development of an Australian anthocyanin food composition database for dietary studies—A systematic process and its challenges. *J. Food Comp. Anal.* 2017, 64, 33–38.
7. Roe, M.A.; Bell, S.; Oseredczuk, M.; Christensen, T.; Westenbrink, S.; Pakkala, H.; Presser, K.; Finglas, P.M. Updated Food Composition Database for Nutrient Intake; EFSA Supporting Publication: Parma, Italy, 2013.
8. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 2016, 3, 160018.
9. Research Data Alliance, GEDE—Group of European Data Experts in RDA. Available online: <https://www.rd-alliance.org/groups/ge-de-group-european-data-experts-rda> (accessed on 20 May 2019).
10. Holden, J.M.; Bhagwat, S.A.; Patterson, K.Y. Development of a Multi-nutrient Data Quality Evaluation System. *J. Food Comp. Anal.* 2002, 15, 339–348.
11. Westenbrink, S.; Roe, M.; Oseredczuk, M.; Castanheira, I.; Finglas, P. EuroFIR quality approach for managing food composition data; where are we in 2014? *Food Chem.* 2016, 193, 63–74.

12. Greenfield, H.; Southgate, D.A.T. Food Composition Data; Production, Management and Use, 2nd ed.; FAO: Rome, Italy, 2003.
13. Probst, Y.C.; Cunningham, J. An overview of the influential developments and stakeholders within the food composition program of Australia. *Trends Food Sci. Technol.* 2015, 42, 173–182.
14. European Commission. Quality Schemes Explained. Available online: https://ec.europa.eu/info/food-farming-fisheries/food-safety-and-quality/certification/quality-labels/quality-schemes-explained_en (accessed on 20 May 2019).
15. European Food Information Resource (EUROFIR). Available online: <http://www.eurofir.org/> (accessed on 20 May 2019).
16. International Network of Food Data Systems (INFOODS). Available online: <http://www.fao.org/infoods/infoods/en/> (accessed on 20 May 2019).
17. UNESCO. Intangible Cultural Heritage, Mediterranean Diet. Available online: <https://ich.unesco.org/en/RL/mediterranean-diet-00884> (accessed on 20 May 2019).
18. Brandbank. Available online: <https://www.brandbank.com/> (accessed on 20 May 2019).
19. Global Trade Item Number, GTIN Info. Available online: <https://www.gtin.info/> (accessed on 20 May 2019).
20. European Commission. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004 Text with EEA relevance. *Off. J.* 2011, 54, 18.
21. European Food Safety Authority. Standard sample description for food and feed. *EFSA J.* 2010, 8, 1457.
22. 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).
23. Slimani, N.; Deharveng, G.; Unwin, I.; Southgate, D.A.T.; Vignat, J.; Skeie, G.; Salvini, S.; Parpinel, M.; Møller, A.; Ireland, J.; et al. The EPIC nutrient database project (ENDB): A first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur. J. Clin. Nutr.* 2007, 61, 1037–1056.
24. Slimani, N.; Deharveng, G.; Unwin, I.; Vignat, J.; Skeie, G.; Salvini, S.; Møller, A.; Ireland, J.; Becker, W.; Southgate, D.A.T.; et al. Standardisation of an European end-user nutrient database for nutritional epidemiology: What can we learn from the EPIC Nutrient Database (ENDB) Project? *Trends Food Sci. Technol.* 2007, 18, 407–419.
25. EuroFIR. Network of Excellence and NEXUS Projects. Available online: <http://www.eurofir.org/our-resources/noe-and-nexus-projects/> (accessed on 20 May 2019).
26. The Royal Society, Global Challenges Research Fund. Available online: <https://royalsociety.org/grants-schemes-awards/grants/gcrf/> (accessed on 20 May 2019).

Retrieved from <https://encyclopedia.pub/entry/history/show/33378>