A network model for improving access to food and nutrition data

DAGOBERT SOERGEL, PhD;
JEAN A. T. PENNINGTON, PhD, RD;
ANNE MCCANN, MLS; JOANNE M. HOLDEN, MS;
ELIZABETH C. SMITH, MLS; ROBERT C. WILEY, PhD

ABSTRACT: In this article we propose a network in which existing food composition and consumption databases are linked through a master database of complete and detailed food descriptions. The proposal arises from an analysis of the importance of food data, their descriptive and analytical nature, and their uses. Lack of detail and standardization in food description hinders the retrieval of food and nutrition data from various databases and the integration of such data. Standardized food descriptions can be developed and maintained in the master database, which can then serve as the interface to the many existing databases of analytical data (especially food composition data) and to databases containing data on food production, consumption, and effects, thereby linking these databases in a coordinated system, or network. The ability to link food-related databases by standardized food descriptions offers a powerful tool for scientists and practitioners in the field of food and nutrition. J Am Diet Assoc. 1992; 92:78-82.

The recognition of the relationship between diet and health, the heightened concern for food safety, the development of new food products and manufacturing methods, and the need to make the best possible use of available resources to ensure an adequate food supply for a growing world population are all problems that call for vastly improved access to food and nutrition data (1-3). Working on these problems are nutritionists, epidemiologists, dietitians, food scientists and engineers, food chemists, economists, marketing specialists, genetic engineers, and others, located in universities, research institutions, government agencies, foodservice facilities, food industry, trade associations, and consumer groups. They use food data to estimate the intake of nutrients and exposure to other substances by various population groups, study the relationship between diet and health, write regulations, monitor compliance with regulations, plan diets and menus, design food products, develop nutrition labeling and claims, monitor the truthfulness of label claims, evaluate the demand for agricultural products, and assess the value of new varieties produced through breeding (4,5).

The descriptive, analytical, and food use data required for these purposes are scattered over many food databases. Access to the data is difficult because each database has its own system for identifying and describing foods. Combining data from two or more databases thus becomes difficult and time consuming.

To facilitate access to food data, we propose a network that will link existing food composition and consumption databases through a master database of complete and detailed food descriptions. In this network, the effort to describe a food will be expended only once; all databases recording data about the food can refer to this description. In searching the network, a user will first consult the master database to retrieve a list of foods meeting the required descriptive characteristics and then request data about these foods from various databases in the network.

The key elements of our proposal are (a) a master database of detailed food descriptions, (b) a flexible, standardized language for the description of foods in the master database, (c) identification of the foods by unique codes, and (d) linkage from the master database to other food and nutrition databases through universal use of the food identification codes.
NATURE AND USES OF FOOD DATA

The organization of food databases must proceed from a thorough understanding of the nature and uses of food data. Our discussion of requirements for food databases and our proposed network rely heavily on the explanations given in this section.

The data pertaining to a food product can be categorized into three types (Table 1):

1. **Descriptive data** deal with the definition of a food and the ingredients and the process used in making, storing, and handling the food.
2. **Analytical data** deal with the properties of the food as determined through quantitative or qualitative analysis.
3. **Food production and use data** deal with production and sales statistics and who consumes how much of the food, where, when, and with what effects.

Many studies investigate the effects of descriptive characteristics (the independent variables) on analytical properties or on the use and effects of foods (the dependent variables), for example, the effects of heat treatment on the shelf life of milk or the effect of variety and soil type on the nutrient content of potatoes. Analytical data may serve in turn as independent variables, as in the use of nutrient composition data for diet analysis and nutrient intake studies and in studies of the effects of nutrient intake on health.

Data on foods are not restricted to data on finished food products; they also include food recipes and food standards of identity. Recipe databases (including cookbooks) are used when foods are prepared in the home or in industry. A database containing standards of identity for food products would be useful to regulatory personnel, food producers, and consumers. Recipes and food standards could be expressed in the same language used for the descriptions of finished foods; this would make it easy to determine whether a food description complies with the stipulations of a standard. Furthermore, once a food is prepared in accordance with a recipe, the recipe is the description of the food; the same is true to a lesser extent for foods prepared in compliance with a standard. (Standards often allow for options, and a precise description of a particular food must reflect the option chosen.)

A fruitful technique of using food data that is particularly dependent on database organization is integration, that is, the retrieval and combination of food data from several databases. Integration may be horizontal or vertical (6).

**Horizontal integration** pools or compares like data from several databases. Examples are pooling data from two or more nutrient composition databases to cover a wider range of foods; comparing nutrient values given in different databases to derive the best value; or using two or more editions of a food consumption survey database for a comparison over time (7,8).

**Vertical integration** correlates different types of data from several databases. For example, correlating data on nutritional and toxic elements in foods with data on food consumption provides estimates of the intake of these elements (7,9); combining data on nutrient composition, sensory evaluation, availability, and cost is helpful in designing a food product that is both nutritious and marketable.

## Requirements for the Description of Foods in Databases

Effectiveness and ease in using food data depend on the accuracy and completeness of food description, both for applications that use food data from a single database (e.g., diet planning) and for applications that combine or integrate data from several databases (10,11). Thus, increasingly sophisticated uses of food data impose requirements for the description of foods in databases. Food descriptions must support two overlapping functions:

- Finding foods and matching foods with precision: (a) searching for foods from many points of view (e.g., plant or animal source, processes applied, preservation method, packaging, or any combination) and analyzing data for the foods found; (b) matching a food at hand with a food for which data are reported in a database; and (c) matching food records in two or more databases to make sure they pertain to the same or sufficiently similar foods

- Finding more information about a food once it has been identified

A food product and the data about it should be accessible through all of its important descriptive characteristics, in-
Table 2
Examples of food databases maintained by government agencies

US Department of Agriculture (USDA) Nutrient Data Base for Standard Reference (14)
Food composition database maintained and published by the USDA's Human Nutrition Information Service. Includes data for proximate components, vitamins, and minerals for more than 4,000 foods. Serves as the basis for many other databases maintained by universities and hospitals.

Total Diet Study (7,20)
Food monitoring database maintained by the Food and Drug Administration (FDA)/Center for Food Safety and Applied Nutrition (CFSAN) since 1961. Current program includes data for 234 frequently consumed foods. For each year there are three files: selected minerals in foods; pesticide residues, industrial chemicals, and heavy metals in foods; and radionuclides in foods. Also includes data on intakes of these substances, based on national food consumption data.

Food Label and Package Survey (FLAPS)
Food labeling information database maintained by the FDA/CFSAN. Label information (e.g., nutrient content, ingredients, dietary claims) collected once every 2 years. The information is maintained and tracked in a computerized system to assess changes in food-labeling practices. Foods are selected based on market share within food groups.

NFCSs provide data on the food and nutrient intake of representative samples of the US population. Use the Nutrient Database for Individual Food Intake Surveys, which is generated by a system using the Primary Nutrient Data Set (based primarily on the USDA Nutrient Data Base for Standard Reference), the Recipe File, and the Nutrient Retention Factor File. CSFII provides data on the dietary, health, and nutritional status of representative samples of the US population.

SIREN (Scientific Information Retrieval and Exchange Network)
Bibliographic index database maintained by the FDA/CFSAN. Includes more than 900,000 citations to petitions, correspondence, speeches, documents, and so forth, relating to FDA regulatory activities concerning foods, food and color additives, and animal drugs.

Table 3
Food name and food code definitions and examples

Food name—A name that designates a food, part of everyday language
Examples:
Chilled spicy tomato soup, also called gazpacho
Sugar-coated puffball wheat, a brand of which is called Tuff

Food code—A code created specifically to identify a food in a database, usually some combination of letters and/or digits
Example: FPS12

Explicit food name or code—A food name or code built up from component terms or codes, respectively, such that each component expresses a characteristic or aspect of the food; together, these components form a capsule description of the food; the meaning of an explicit food name or code can be inferred from its components
Examples:
Explicit food names: chilled spicy tomato soup
Explicit food code: TY157.S0123.PT093.PR076.PR123.FP368

Nonexplicit food name or code—A food name or code that is not constructed from components that express characteristics of the food; the meaning of a nonexplicit name or code cannot be inferred; it must be known or looked up in a reference source or database
Examples:
Nonexplicit food names: gazpacho, Tuff
Nonexplicit food code: FPS12; the Universal Product Code for a food

Including animal or plant source and part used, the processing of the food (raw or cooked, cooking method used, other treatments applied, preservation), packaging, and other descriptive characteristics known to affect nutrient content and consumption patterns. For example, a nutrient database that can be searched not just for potato but more specifically for mashed potato allows for more accurate dietary analysis or planning. Another example is a study of the shelf life of lactic-acid-fermented plant products, which would profit from retrieving all (or most) such products and data about them.

A user who needs data applicable to a food at hand, such as a store-bought product or a local fruit or vegetable variety, must have access to sufficient descriptive data about the foods in the database to ensure that the food at hand is identical, or at least comparable, to the food found in the database; otherwise, the analytical values given in the database may not be valid when applied.

The comparison or combination of data from two or more databases requires food descriptions that manifest the correspondence between the food products in these databases so that the identical or similar nature of two foods can be determined. A researcher must be able to determine whether two food descriptions in different databases refer to the same food or to foods that agree with respect to descriptive characteristics of importance for a given purpose (12). In studies that combine food composition data with consumption data to assess the intake of nutrients and other substances, imprecise descriptions of the foods analyzed as well as the foods consumed may lead to large margins of error. For example, the levels of specific carotenoids in vegetables are affected by the cooking method (13). If the cooking method is omitted from the food descriptions, errors may result.

To interpret analytical and food use data and apply them correctly in these and other situations, a user requires quite detailed descriptive data about foods found in a database. Interpretation of nutrient data requires knowledge of preparation (e.g., raw or cooked, boiled or fried), details such as cut of meat, and in some cases length of storage. For example, assessing the safety of a low-acid or acidified low-acid canned food requires detailed data on the time, temperature, and other aspects of processing.

Changes in the food marketplace must also be considered. Breeding and new feeding methods have produced meat with lower fat content; meat cuts are marketed well trimmed; and products of increased complexity are continually being introduced.

Accurate food description is required to capture the detailed data on nutrient composition (e.g., nutrient data on different cuts of beef at different degrees of cooking) produced by laboratories using improved methods for analyzing foods. It is also required to capture the more precise data on what people eat that are provided by consumption surveys using sophisticated data collection techniques.

Other issues in using food data, such as the quality of analytical data and the compatibility of the laboratory methods and computations used to arrive at these data, are beyond the scope of this article.

PROBLEMS IN USING EXISTING DATABASES
The US Department of Agriculture (USDA) maintains comprehensive databases on food composition and consumption (14). The National Center for Health Statistics (NCHS) develops databases concerning health and nutritional status (15). The 1988 Nutrient Data Bank Directory (16) lists 112 additional food composition databases in the United States and Canada, some of which are made available by universities.
and private organizations. Many of these databases are derived from USDA nutrient databases with added data from the food industry, the scientific literature, and other sources. Smaller nutrition databases are available for use on personal computers (17). Further sources of nutrient data and databases are listed in references 18 and 19. Examples of food databases maintained by government agencies are given in Table 2. Other databases are maintained for exclusive use by many food companies, universities, and government agencies.

Problems in using these databases stem from nonstandardized food identifiers (food names and food codes; see Table 3) and nonstandardized food descriptions that lack the detail needed for many purposes. Most databases identify foods through short codes; unfortunately, these codes are often not the same from one database to another, or even from one edition of a database to the next, because the number of foods and the foods themselves change over time. Most databases describe foods through names consisting of several components, each of which refers to a food product characteristic (eg, cheeseburger with condiments; corn with butter; beer, regular). The information given in such descriptions varies; although it is satisfactory for some purposes, it is not sufficient for others. In particular, the information rarely allows the user to determine whether two food descriptions in different databases refer to the same food.

Furthermore, descriptive characteristics are seldom explicit enough to be manipulated by computer for retrieval or comparison of databases. Some databases provide access to foods only through an arrangement into food groups (eg, legumes, fats and oils, breakfast cereals). However, such classifications are inflexible, allowing access from only one point of view, and they do not do justice to the complexity of food description (12). For example, a search for products made from soybeans might find the soybeans under legumes but miss the soy formula under infant foods, the textured vegetable protein under meat analogues, or the soy oil under fats and oils.

As with food identification codes, there is no common standard for a method of food description across databases. This lack of standardization makes the creation, maintenance, and use of databases unnecessarily difficult and expensive. Database producers expend effort on constructing separate schemes for food description and even more effort applying these schemes to describe hundreds or thousands of products. Users must understand the differing structures of several databases and can rarely retrieve, compare, or combine data with any degree of confidence or ease.

Several efforts are under way to find and implement solutions to these problems. The USDA has been working to improve the food names used in Revised Agriculture Handbook No. 8 and in its Nutrient Data Base for Standard Reference (14) to provide more precise descriptions of the foods. Although in the past nationwide food consumption surveys conducted by the USDA and NCHS have used different food names, surveys by both agencies presently use the same food names and descriptions, which will make comparison of consumption data from the USDA and NCHS surveys much easier. The International Network of Food Data Systems (INFOODS) (27-30) and EUROFOODS (31) have been working on methods for food description on an international scale, dealing with the additional problems caused by differences in language and culture as well as differences in food sources and processing. Finally, the Center for Food Safety and Applied Nutrition at the Food and Drug Administration has developed and implemented a very flexi-

![Diagram](https://via.placeholder.com/150)

**Figure 1.** A system model for a food database network. **SIREN = Scientific Information Retrieval and Exchange Network, a system for the storage and retrieval of external and internal documents.**

ble language for the description of foods, the Factored Food Vocabulary (now called Langual, for Lange Alimentaire [32]), which has been tested for international use with translations into French and Hungarian.

### A Model for a System of Food Databases

We propose a comprehensive solution that exploits the possibilities offered by modern computer and communication technology. Our model addresses the problems of access to food data by separating descriptive data from analytical and food use data.

Descriptive data on the one hand and analytical, food production, and use data on the other differ in their nature and in the degree of consensus that can be achieved about them. The descriptive characteristics of a food can be ascertained and expressed with reasonable ease and objectivity, provided that the information needed for a precise description is accessible. Thus, consensus about food descriptions that can be used by all concerned is possible. On the other hand, in nutrition analysis and surveys of consumption and health effects, methodological issues and judgments play a significant role. There is much more room for disagreement in the interpretation and validation of analytical and food use data than in the area of food description. Because descriptive data differ in nature from analytical and food use data, they require different handling. Descriptive data are amenable to efficient centralized efforts, whereas analytical and food use data require local attention and control. Yet it is desirable that all data about a food product be linked.

Fortunately, computer and communication technology provide a solution. Linkage does not require that these data all reside in the same database. Food descriptions can be stored in a master database accessible to all. Linkage to the many databases that contain analytical and food use data can be achieved by assigning to the food product a standardized code that is brief, unique, and permanent. This code then identifies the food throughout all databases—in the master database of food descriptions as well as in the many databases of analytical or food use data (Figure 1). Descriptions can also be copied from the master database into a local database, and local codes for foods can coexist with the standardized codes. Either way, a food product needs to be described only once, and one precise description is less costly and more
useful than many independent, superficial, and inconsistent descriptions.

The master database would cover all food products included in any participating database on various levels of specificity, from product types to generic products to brand name products or recipe items. Each food product would be fully described using a standardized food description language, which would provide the necessary unified structure. (The nature of such a language will be discussed in a separate article.) As this language uses a building-block approach to the construction of food descriptions, the effort required for food description decreases as the database grows.

**IMPLICATIONS**

The detailed food descriptions in the master database would enable the user to retrieve food products from many points of view, using appropriate descriptive characteristics. Each food product would be identified by its standard code. With these codes the user could then access analytical and food use databases, thereby opening up new possibilities for comparing and correlating different types of data and performing data analyses. For example, the user could easily find all nutrition databases that include a given food and compare nutrient values. (Although this article deals neither with compatibility of nutrient definitions and analytical methods nor with conversion of measuring units, these problems could be addressed in the framework of our proposal.) The user could also look for a correlation between any food characteristic and consumption patterns. Conversely, when examining a computer screen display of analytical or food use data, the user could access the master database to call up detailed descriptions of the foods analyzed. The user could also retrieve a food by its brand name or by its Universal Product Code, possibly with a bar code reader.

Implementation of this proposal would place a powerful tool in the hands of scientists and practitioners in the field of food and nutrition.

**References**


6. Rand WM, Pennington JAT, Murphy SP, Kleslin JC. Compiling Data for Food Composition Databases. The United Nations University. In press. (especially chp 4, data from other sources.)


