

FDA's Factored Food Vocabulary for food product description

Anne McCann, M.S.L.S.,
Jean A.T. Pennington, Ph.D., R.D.,
Elizabeth C. Smith, M.S.L.S., Joanne M. Holden, M.S.,
Dagobert Soergel, Ph.D., and Robert C. Wiley, Ph.D.
*Divisions of Information Resources Management and
Nutrition, Center for Food Safety and Applied Nutrition,
Food and Drug Administration, Washington, D.C.; Nutri-
ent Composition Laboratory, Beltsville Human Nutrition
Research Center, Agricultural Research Service, U.S.
Department of Agriculture, Beltsville, Maryland; and
College of Library and Information Services and Food
Science Program, University of Maryland, College Park*

Abstract The Food and Drug Administration's Factored Food Vocabulary (FFV) uses standardized language to describe characteristics of food products that are important for food safety and nutritional quality. Each food product is described by a set of descriptors from the following factors: product type; food source; part of plant or animal; physical state, shape, or form; degree of preparation; cooking method; treatment applied; preservation method; packing medium; container or wrapping; food contact surface; and user group. The purpose of the vocabulary is to facilitate retrieval of food composition, food consumption, food contamination, and other food-related data relative to these factors. The major advantages of this system are flexibility with specificity, enhanced searchability with economy, and ease of change and updating. *J Am Diet Assoc 88:336, 1988.*

The Factored Food Vocabulary (FFV) of the Center for Food Safety and Applied Nutrition (CFSAN) of the Food and Drug Administration (FDA) provides a standardized language for describing foods, specifically for classifying food products for information retrieval. Designed to serve CFSAN needs, the vocabulary describes those characteristics that affect the safety and/or nutritional quality of food products. Its structure is based on two main ideas: (a) A food product can be described by a combination of several characteristics, each of which may serve as a retrieval term, or descriptor, for the food product. (b) The characteristics can be brought together in a meaningful classification relating them to one another.

Table 1 shows the 12 characteristics (factors) used in FFV to describe a food product. Factor terms are combined to produce a multifaceted description of a product, as illustrated for corn flakes in Table 2. A data base utilizing such food descriptions can be searched for individual factor terms or for combinations of terms to retrieve food products that have been indexed according to their characteristics (e.g., all frozen foods, all canned foods containing mushrooms, all freeze-dried foods with added colors and wrapped in cellophane).

Background

The mission of CFSAN is to ensure the safety and nutritional adequacy of the American food supply. In part, this mission is accomplished by:

- Monitoring the levels of potentially harmful substances in foods, whether present naturally or resulting from agricultural and manufacturing practices or contamination
- Ensuring that direct and indirect food additives are safe and are present at levels no higher than needed to accomplish the intended technical effect
- Assessing the effects of food processing on composition, nutritional quality, and safety of foods
- Preventing food fraud and consumer deception
- Assessing the nutritional status of the United States population and population subgroups
- Evaluating the relationships of diet to health and disease

To obtain information about these topics, CFSAN depends on the use of food composition and food consumption data bases. Use of these data bases has revealed problems in retrieving information that stem from the way they are organized. Traditionally, the foods are listed in data bases alphabetically or by food group (e.g., dairy, meat, grain, vegetable). Such organizations are logical for menu planning and diet evaluation; however, many of the questions that CFSAN seeks to answer cannot be addressed efficiently with such organizations. Also, in most cases, the descriptions do not provide enough information about ingredients, processing, preservation methods, and container materials.

CFSAN often requires information on the number of foods or intake levels of foods with commonalities of food source, packaging, preservation, or cooking method

Table 1. Factored Food Vocabulary (FFV) factors and examples of factor terms

factor notation	factor name	example factor terms
A	product type	dairy product, poultry/poultry product, beverage, gravy or sauce, sweetener
B1	food source	cattle, abalone, wheat, carob bean, garlic, plantain, yeast
B2	part of plant or animal	leaf, fruit, skeletal meat, organ meat, egg, milk
C	physical state, shape or form	liquid; semiliquid; solid; whole, natural shape; divided into pieces
D1	degree of preparation	fully cooked; partially cooked; uncooked, raw
D2	cooking method	sauteed, baked or roasted, griddled, toasted, popped, deep-fried
D3	treatment applied*	enriched, sweetened, egg added, fat or oil added, fat removed, decaffeinated
D4	preservation method	dehydrated or dried, frozen, preserved by adding chemicals
E1	packing medium	packed in broth, packed in gelatin, packed in gravy or sauce, packed in olive oil
E2	container or wrapping	paperboard tray with wrapper; plastic boil-in-bag; glass container, aluminum lid, plastic lining
E3	food contact surface*	ceramic, cork, paperboard, glass, metal, plastic
F	user group*	human food, low fat; human food, sodium-free; human food, very low sodium; human food, reduced calorie

*Multiple values may be used as appropriate to describe a single food.

(e.g., canned foods, freeze-dried foods, foods with artificial colors, foods with aluminum containers, smoked foods). Before FFV was developed, such information could be obtained only by time-consuming manual searches of the printed copy. For example, estimating soy consumption requires retrieval of all products containing soy. This cannot be accomplished by looking only under "soy" in a data base arranged alphabetically by food name or only under "legumes" in a data base arranged according to a classification of food groups (Example 1). The entire data base would have to be searched manually to retrieve soy-based infant formula, textured vegetable protein, flour from soy, meat analogues prepared from soy, and soybean oils and margarines. Besides being tedious, these manual searches may not be successful because the foods are often inadequately described.

The solution to these retrieval problems is to provide more complete food descriptions and to allow for organizations of the data base other than by alphabetic or food group listings. FFV has been developed in response to this need; it allows food products with common characteristics to be retrieved. This effort, which began in 1975, is carried out by the FFV Committee (an interdisciplinary group with expertise in nutrition, food technology, and information/computer science). The committee meets on an ongoing basis to update, review, and improve the vocabulary.

Factors and factor terms

The sequence of the factors shown in Table 1 was chosen to facilitate the writing and reading of food product descriptions and to aid in the comprehension of the vocabulary. After the food product has been characterized overall (factors A, B1, B2, and C), the factor sequence follows stages of processing (D1-D4), packing and packaging (E1-E3), and consumption (F). The factors are designed to answer specific questions:

- A To what food group does the product belong?
- B1-B2 What is the origin of the food product?

Table 2. Example of corn flakes* factored according to FFV

factor notation	factor name	factor term	code
A	product type	breakfast cereal	A258
B1	food source	field corn	B1379
B2	part of plant or animal	seed or kernel, skin removed, germ removed (endosperm)	C208
C	physical state, shape, or form	whole, shape achieved by forming, thickness less than 0.3 cm	E153
D1	degree of preparation	fully cooked	F14
D2	cooking method	cooked by dry heat	G004
D3	treatment applied	flavoring or spice extract or concentrate added	H100
		water removed	H138
		sucrose added	H158
		vitamin added	H163
		iron added	H181
		fortified	H199
D4	preservation method	dehydrated or dried	J116
E1	packing medium	no packing medium used	K03
E2	container or wrapping	paperboard with paper liner	M148
E3	food contact surface	paper	N39
F	user group	human food, no age specification, regular diet	P24

*Food product description: corn flakes, fortified, in cardboard box.

- C What are the physical characteristics of the food product?
- D1-D4 What processing operations have been performed to transform the food or ingredients into the final product?
- E1-E3 What is the product's packing medium? How is the product packaged?
- F Who uses the food product? What are its special dietary characteristics?

The "product type" refers to what would traditionally be called a food group. It is the focal point in defining a food product in terms of the common consumption, functional, and/or manufacturing characteristics of a food.

The "food source" and "part of plant or animal" indicate the origin of the food. The food source is usually the plant (wheat, oat, carrot) or animal (cattle, swine, chicken) from which the food was derived. For multi-ingredient food items, the food source of the main ingredient by weight (other than water) is indexed; other ingredients are included as additions under the factor term "treatment applied." For example, the food source for raisin bread is wheat, and under treatment applied the factor term "fruit added" is listed. The "part of plant or animal" lists such descriptors as leaf, fruit, skeletal meat, organ meat, egg, and milk.

The "physical state, shape, or form" is important for food safety because it affects heat transfer through the product, susceptibility to permeation by chemical substances, and invasion by microorganisms. These physical characteristics may be inherent or may result from processing.

"Degree of preparation" refers to the amount of heat applied to a food product, e.g., "fully cooked," "partially cooked," or "uncooked, raw." "Cooking method" refers to the process by which heat or microwaves are applied to a food product. "Treatment applied" describes processing steps and usually involves adding, substituting, or removing food components or modifying the food or components (e.g., fermentation). Multiple factor terms may be used for a single food product. "Preservation method" is the primary method used to prevent microbial and enzymatic spoilage.

"Packing medium," "container or wrapping," and "food contact surface" are all components of packing and packaging. Packing medium is the substance (within the container) in which the food is packed; it can usually be separated from the food (e.g., the brine in canned vegetables or the syrup in canned fruit). Container or wrapping refers to the main container material, the container form, and the material of the liner, lids, and ends. Food contact surface indicates the container materials that are in direct contact with the food. Multiple values can be assigned for this factor.

"User group" refers to the consumer group for which the product is produced and marketed, e.g., "infant or toddler food, no age specification, regular diet," and "human food, low sodium."

Because of its flexible structure, the vocabulary can be amended by altering definitions, by including or deleting factor terms, or even by adding or deleting entire factors. As new foods enter the marketplace or new techniques are developed, factor terms can be added where appropriate to permit the description of these foods and processes.

Example 1. Excerpt from a hypothetical traditional food classification

baby foods
infant formulas made from soy

breakfast cereals
corn flakes

fats and oils
corn oil
soybean oil

fruits
grapes

grain and starch products
corn flour or meal
soy flour

legumes
soybeans

meat analogs
textured vegetable protein
Note: usually made from soy

Example 2. Excerpts from Factored Food Vocabulary (FFV) hierarchies

product type
grain or starch product
milled grain or starch product
prepared grain or starch product
bakery product
bakery product, sweetened
cake
bakery product, unsweetened
bread
breakfast cereal
macaroni or noodle product

food source
grain or seed producing plant
corn
field corn
field corn, white
field corn, yellow
popcorn
popcorn, white
popcorn, yellow
vegetable producing plant
sweet corn
sweet corn, white
sweet corn, yellow

part of plant or animal
part of plant
seed or kernel
seed or kernel, skin (bran) present
seed or kernel, skin present, germ present
seed or kernel, skin present, germ removed
seed or kernel, skin removed
seed or kernel, skin removed, germ present
seed or kernel, skin removed, germ removed (endosperm)

treatment applied
ingredient added
nutrient or dietary supplement added
vitamin added
vitamin A added
vitamin C added
vitamin D added
food modified
physically/chemically modified
bleached
puffed
textured

Hierarchical structure

To facilitate retrieval and aggregation, the descriptors within each factor are arrayed in a hierarchy from broader to narrower terms (Example 2). For example, one might search specifically for "vitamin A added," more broadly for "vitamin added," or still more broadly for "nutrient or dietary supplement added," or one might aggregate consumption of all "prepared grain or starch products." The hierarchical arrangement also displays the vocabulary in a logical way to facilitate indexing and retrieval. The systematic presentation of all possible factor terms for a specific factor at one time permits a more standardized approach to indexing; it allows the indexer to select the most specific term that can be assigned on the basis of the information at hand. This leads to improved search results with respect to both retrieving relevant items and excluding irrelevant ones.

Scope notes

The success of the vocabulary as an instrument in the exchange of data depends on the degree to which the factor terms are understood and used consistently by coders, indexers, and searchers. To forestall the tendency of different users to assign different meanings to the descriptors, the vocabulary includes many scope notes that define the terms and explain when and in what contexts they should be used. Often scope notes refer to more specific definitions, such as those in the Code of Federal Regulations (CFR) (1). A few FFV scope notes are shown in Example 3.

CFSAN thesaurus and FFV

FFV has been developed as an information retrieval thesaurus, using standard conventions for displaying terms and cross references. FFV is a subset of an even larger, fully computerized thesaurus maintained for all of CFSAN's information retrieval systems. Many of the factor terms have synonyms, which are included in the thesaurus to direct indexers and searchers to the descriptors. The scientific names for many food sources are also included in the thesaurus. "Use" and "used for" cross references show these relationships. In addition to "use" and "used for," the thesaurus displays cross references to broader terms, narrower terms, and related terms, as well as "code" and "code for," which parallel "use" and "used for." (See Example 4.)

In terms of data processing, FFV has the following components:

- A computer-based thesaurus (dictionary), which is an adjunct to CFSAN's information retrieval systems. Seventeen subfiles (microthesauri) contain the descriptors, synonyms, codes, and so-called pointers that generate cross references between descriptors. An additional subfile houses food names with their factor codes and external codes identifying the same foods in other data bases (Table 3).
- Software for assembling, formatting, and printing the thesaurus, which uses Computer Corporation of America's Model 204¹ online data base management and Informatics Mark IV² file management systems.

Example 3. FFV scope notes

brown sugar—soft sugar whose crystals are covered by a film of refined dark syrup that imparts color, flavor, and moisture.

cultured milk product—milk modified by adding acid-producing and/or flavor-producing bacteria under controlled conditions.

packed in fruit juice—packed in unsweetened, single-strength juice or juice prepared from concentrated fruit juice and water; if sweetener is added to the fruit juice, use the appropriate term under "packed in sweetened liquid."

sucrose added—used when sucrose, specifically cane sugar or beet sugar, is the second or third ingredient in order of predominance or if the sucrose content is known to be 5% or more. Note that on a product label or in a recipe, "sugar" means "sucrose."

Example 4. Cross references from the FFV thesaurus

sodium chloride added	use (UT)	salted
millet	use for (UF)	panicum miliaceum
wine	broader term (BT)	alcoholic beverage
component substituted	narrower term (NT)	fat substituted
nut or edible seed	related term (RT)	plant used for producing fat or oil
uncooked, raw B1170	code (CDS) code for (CF)	F03 black olive

Storage and processing of food description records

For efficient internal storage and for information processing, factor terms are assigned nonhierarchical alphanumeric codes. The codes consist of a letter for each factor, followed by a number assigned sequentially as terms are introduced within the factor. They bear no relationship to the FFV factor notation. Alphanumeric codes for the factor terms assigned to cornflakes can be seen in Table 2.

The record for a food product contains the codes for the assigned factor values and can be retrieved under each of these individually or in combination. At present, indexers enter factor codes, and searchers search by factor codes. For example, to retrieve "breakfast cereal," one searches for A258, and to retrieve "sucrose added," one searches for H158. A more advanced user interface will accept factor terms.

A food description record may be linked to records for the same food in one or more other data bases (such as the FDA's Total Diet Study or a food consumption survey), by including an identifying external code (such as TDS-153). This linkage makes the power of FFV available for searching each of these data bases and thus provides a common denominator for searching multiple data bases.

Displaying and searching

The vocabulary can be displayed in a variety of formats, as appropriate for specific purposes. Those most commonly used are hierarchical displays of the factor terms and alphabetic displays of the factor terms with synonyms. A data base coded with FFV can be sorted and displayed by any factor, e.g., by product type to get a listing of food groups, by treatment applied, by preservation method, by container or wrapping, or by any of the other FFV factors. Example 5 shows foods from the FDA Total Diet Study (2) displayed by the various factors. Any one of the approximately 1,750 FFV factor terms or various combinations of two or more may be searched to locate foods with given

¹Computer Corporation of America, Cambridge, MD.

²Informatics; now Sterling Software, Canoga Park, CA.

Table 3. Sample records from the file of food names

food no.	food name	FFV code	external codes for other data bases*				UPC
			TDS	NFCS	USDA	USDA (rev.)	
F001	corn flakes	A258.B1379.C208. E153.F14.G004.H100. H138.H158.H163.H181. H199.J116.K03.M148. N39.P24	071	5730100	866	†	†
F002	Kellogg's corn flakes	A258.B1379.C208. E153.F14.G003.H100. H138.H158.H181.H199. H213.H214.H215.H216. J116.K03.M148.N39.P24	†	5730100	†	08020	381100
F003	buttermilk, fluid	A101.B1201.C235. E114.F14.H101. J135.K03.M001.N01. P24	005	1111500	509	01088	†

*TDS=Total Diet Study; NFCS=Nationwide Food Consumption Survey, 1977-78, U.S. Department of Agriculture; USDA=U.S. Department of Agriculture Handbook No. 8, 1963; USDA (rev.)=U.S. Department of Agriculture revised Handbook No. 8, 1974-1986; UPC=Universal Product Code.

† Not included in this data base.

Example 5. Total Diet Study foods displayed by FFV factors**product type listing**

alcoholic beverage
wine, table
bread
bread, rye
bread, white
cornbread, homemade
breakfast cereal
corn flakes in a box
shredded wheat in a box
corn grits, enriched, cooked
oatmeal, cooked

food source listing

cabbage
cabbage, boiled from raw
coleslaw with dressing, homemade
sauerkraut, canned
chicken
chicken, fried, homemade
egg, scrambled with milk and fat
egg, soft boiled
pot pie, chicken, frozen
field corn
corn chips
corn flakes
corn grits, enriched, cooked

container or wrapping listing

aluminum tray or pan, cardboard cover
pie, apple, commercial, frozen
paperboard container
cake, yellow from mix
macaroni and cheese from box mix
corn flakes in a box
shredded wheat in a box
glass container
peanut butter
wine, table

factors and then by major categories and subcategories, down to the most specific level) provides a convenient way for the searcher to review all the factor terms to determine which ones are needed for the query at hand. In the alphabetic thesaurus, one can look under synonyms and find the corresponding descriptors; one can also find related terms. Familiarity with the entire vocabulary is needed in order to assure that no factor terms important for a search are overlooked.

Uses of the vocabulary

To try out the vocabulary, CFSAN has assigned FFV descriptors to 653 foods from three data bases: all 234 foods from FDA's Total Diet Study (2), 228 foods (5%) from the food composition data base of the U.S. Department of Agriculture's 1977-78 Nationwide Food Consumption Survey (3), and 191 foods (5%) from the U.S. Department of Agriculture's Nutrient Data Base for Standard Reference (4). Trial searches have sorted the Total Diet Study foods for added sodium, added sugar, canned foods, enriched foods, and so forth. Coding time for most foods ranged from less than 1 minute to several minutes, with multi-ingredient foods taking longest. The vocabulary is currently used to index food products in the Center-wide information system for petitions and other documents.

In a collaborative effort between FDA and the National Cancer Institute (NCI), the foods in the Nutrient Data Base for Standard Reference (augmented by CFSAN to add sections from Release 5, 1985) have been factored according to the FFV. NCI will use the factored data base in its attempt to match foods and nutrient values among various data bases. The nutrients of concern will be those thought to have some relationship to cancer (e.g., dietary fiber, carotenes, selenium). This data base will provide CFSAN an opportunity to study the effectiveness of the vocabulary in information retrieval and aggregation.

The ultimate goal of the FFV for FDA is to be able to provide quick and comprehensive responses to questions of public health significance concerning specifically defined foods. The needs of FDA for food composition and

characteristics. This system overcomes the problem of manual searching for descriptors that cross food group lines. The retrieval process is therefore accurate and efficient.

The food vocabulary's hierarchical arrangement (first by

food consumption data cannot be easily predicted; however, the FFV allows for specific searches for a wide spectrum of descriptors, such as canned foods containing mushrooms, enriched foods containing corn, or partially cooked foods with plastic wrap.

One potential use of the FFV is to match foods between various coded data bases. To link and exchange data among various food-related data bases, the name and descriptive terms for the foods, nutrients, and other food constituents found in these data bases must be specific, easily understood, and unambiguous. Efficient storage, retrieval, and processing of data in computerized food-related data bases require a controlled vocabulary with a glossary or thesaurus that clearly defines the vocabulary terms. If one had a coded food in one data base and wanted to find the closest match to this food in one or more other coded data bases, the system could be asked to retrieve all foods with all identical factor terms, foods with 11 of 12 factor terms, foods with 10 of 12 factor terms, and so forth. Such a procedure could also help to identify and find nutrient information about products whose common names are not easily translated from their original language.

Other systems using the factoring approach

Several other computerized systems use a factored vocabulary for independently considering the characteristics of foods or feeds. The International Network of Feed Information Centers (INFIC) has developed a databank system that uses a faceted vocabulary system to describe six characteristics of feeds (origin, part, process, growth stage, cut, and grade) (5). The descriptive information and numerical data associated with the feeds can be stored, summarized, retrieved, and printed in various formats. Approximately 21,000 feeds have been described by the INFIC system, and each of these is registered with a five-digit International Food Number (IFN). The IFNs are used for data processing and selection and for linking information across data files. The thesaurus contains approximately 5,000 descriptors, each given in English, French, and German.

The Australian Department of Health uses a food factoring system that includes factors and codes for food group, food name, physical state, process, portion analyzed, cooking method, additives, brand, and size (6).

The International Network of Food Data Systems (INFOODS), was formed at an international meeting in Italy in 1983 to promote and facilitate the development and exchange of food composition data that are complete, reliable, and current (7). INFOODS is developing an open-ended, multifaceted food description system for international use. This system will be based on the same general principles as the INFIC system and FFV. Like them, the INFOODS terminology system will be

flexible and responsive to computer technology for data entry, summarization, and statistical manipulations and for user-responsive retrieval.

Advantages of the factoring approach

The factoring approach was selected by CFSAN to achieve flexibility with specificity in searching for information about foods. Identification of classifying concepts (factors) and their separation into elements (factor terms) permit the elements to be used individually for retrieval whenever appropriate or to be omitted if not needed. Indexing by factor terms makes it possible to retrieve products having a specified food source, whereas the food sources are widely separated when the products are indexed according to food groups (see the soybean example described in the "Background" section). It is also possible to freely combine characteristics in searching, choosing the appropriate level of specificity for each, whereas the traditional approach to food classification is rigid, allowing only combinations expressly included.

System efficiency is heightened by storing each factor value only once in the vocabulary. Yet the potential number of different combinations is greatly expanded because any factor term can be used with any other appropriate factor term to describe a food product. The vocabulary is maintained as a computerized file that can be updated readily. Changes made on the basis of new information or to correct errors are available to indexers speedily, but they do not alter any product names or codes assigned in other files.

As a consequence, FFV will enable faster and more precise retrieval of food-related data from CFSAN's information files. A broader objective, which is shared with users of data bases worldwide, is the standardization of food product descriptions and designations for use in both CFSAN and external data bases. The authors are considering the general principles of a food description language that can be used to link data from many data bases. This will be the subject of a later report.

References

- (1) Code of Federal Regulations: Title 21—Food and Drugs. Parts 100-199. Washington, DC: Government Printing Office, 1986.
- (2) Pennington, J.A.T.: Revision of the Total Diet Study food list and diets. *J Am Diet Assoc* 82:166, 1983.
- (3) U.S. Department of Agriculture: Nutrient Data Base for Individual Food Intake Surveys: Release 1. Data tape. Springfield, VA: National Technical Information Service, 1980.
- (4) U.S. Department of Agriculture: Nutrient Data Base for Standard Reference: Release 2. Data tape. Springfield, VA: National Technical Information Service, April 1982.
- (5) Harris, L.E., Haendler, H., Riviere, R., and Rechaussat, L.: International Feed Databank System: An Introduction into the System with Instructions for Describing Feeds and Recording Data. International Network of Feed Information Centers Publ. No. 2. (Prepared on behalf of INFIC by the International Feedstuffs Institute.) Logan, UT: Agricultural Experiment Station, Utah State University, 1980.
- (6) Cashel, K., Australian Department of Health. Canberra, Australia: Personal communication, July 1986.
- (7) Rand, W.M.: Food composition data: Problems and plans. *J Am Diet Assoc* 85: 1081, 1985.