USING A FACTORED VOCABULARY FOR FOOD DESCRIPTION AND NUTRIENT COMPOSITION DATA RETRIEVAL: The FDA's Factored Food Vocabulary

Elizabeth C. Smith, M.S.L.S.

Food and Drug Administration
Center for Food Safety and Applied Nutrition
Washington, D.C. 20204

My intention in this slide/computer presentation is to explain the organization of the FDA's Factored Food Vocabulary, known by its acronym FFV, and to show how it is being used in an information retrieval system at the Center for Food Safety and Applied Nutrition (CFSAN). This system is the Food Monitoring Database. For those of you who would like to learn more about the FFV, I recommend a recent paper entitled "FDA's Factored Food Vocabulary for Food Product Description' which appeared in the March issue of the Journal of the American Dietetic Association. This paper was written by members of the FFV Committee, who meet regularly to update and develop the vocabulary.

The Factored Food Vocabulary is a standardized vocabulary for the description of food products in databases. It provides flexibility with specificity. Its structure is based on two main ideas: first, that a food product can be described by a combination of several characteristics, each of which may serve as a retrieval term, and second, that these characteristics can be brought together in a meaningful classification that relates each characteristic to the others.

The FFV is composed of 13 viewpoints or FACTORS. The factors are arranged so that certain factors with related viewpoints are grouped together. For example, FOOD SOURCE and PART OF PLANT OR ANIMAL deal with the origin of the food. DEGREE OF PREPARATION, TREATMENT APPLIED, and PRESERVATION METHOD concern processing operations that modify the food product. PACKING METHOD, CONTAINER OR WRAPPING, and FOOD CONTACT SURFACE concern the packing and packaging of the food.

Each factor contains a number of FACTOR TERMS or DESCRIPTORS, one or several of which may be selected to describe a food product. With most factors, a single descriptor is selected; however, multiple factor terms may be employed in TREATMENT APPLIED, in FOOD CONTACT SURFACE, and for special dietary characteristics found in CONSUMER GROUP/DIETARY USE. Each factor is applied independently of the others. A food product such as HOMEMADE BREADED FRIED CHICKEN is described as follows:

FACTOR	FACTOR TERM
PRODUCT TYPE	Poultry or poultry product
FOOD SOURCE	Chicken
PART OF PLANT OR ANIMAL	Skeletal meat part, with bone, with skin
PHYSICAL STATE, SHAPE OR FORM	Whole, natural shape
DEGREE OF PREPARATION	Fully cooked
COOKING METHOD	Cooked with added fat
TREATMENT APPLIED	Fat or oil-coated
	Grain added
	Breaded or batter-coated
PRESERVATION METHOD	No preservation method used
PACKING MEDIUM	No packing medium used
CONTAINER OR WRAPPING	No container or wrapping used
FOOD CONTACT SURFACE	No food contact surface present
CONSUMER GROUP/DIETARY USE	Human food, no age specification, regular diet

Factor terms may be augmented in several ways. Many factor terms have definitions describing what they are, what they include or exclude, or how they are intended to be used within the vocabulary. Many definitions specify to what level within a label or recipe an added ingredient is indexed. For example, MUSHROOM ADDED should be used whenever "mushroom is present in a recipe or appears at any level in label ingredients, whereas DAIRY PRODUCT ADDED should be used only when the indexed dairy product is the second, third, or fourth ingredient.

Numerous synonyms point back to factor terms. These enable the user to identify an organism by both its common and its scientific name, or to locate vernacular terms, such as BROWN AND SERVE.

Finally, each set of terms for a given factor is arranged in a hierarchy from broader to narrower terms. For example, GRAPEFRUIT, LEMON, and ORANGE are hierarchically indented under CITRUS FRUIT. Types of oranges, such as SEVILLE ORANGE, are indented under ORANGE. Mixture terms such as GRAPEFRUIT AND ORANGE are indented under both GRAPEFRUIT and ORANGE. This arrangement provides a meaningful classification that relates terms within a factor to each other. Hierarchy also provides a means of relating the factors themselves to each other. The hierarchical arrangement for DAIRY PRODUCT, found in the PRODUCT TYPE factor, parallels the hierarchical arrangement of DAIRY PRODUCT ADDED, found in the TREATMENT APPLIED factor.

Where does this lead? Once the food names in a database have been described by the FFV, they may be linked to other similar or identical food names in other databases for retrieval of analytical, consumption, or bibliographic information.

I will now show you how information may be searched and displayed on-line. To conserve time, I have prepared in advance a diskette that illustrates the vocabulary - first we'll see the CFSAN THESAURUS, which arrays the vocabulary, and then we'll see CFSAN'S FOOD MONITORING DATABASE, which uses the vocabulary as a link to food names in databases.

CFSAN THESAURUS

The entire scope of the vocabulary can be demonstrated by showing all 13 factors. PRODUCT TYPE, FOOD SOURCE, TREATMENT APPLIED, and CONSUMER GROUP/ DIETARY USE are some examples.

The PRODUCT TYPE factor is selected and displayed hierarchically. Some of the broad categories of factor terms include DAIRY PRODUCT, FRUIT OR VEGETABLE PRODUCT, GRAIN OR STARCH PRODUCT, and PREPARED FOOD PRODUCT. The alphanumeric code that follows each term is the means by which the term is stored in the computer and linked to a food product.

The user may display part of a hierarchy for a factor without having to see all the other terms that appear in the factor. FRUIT OR VEGETABLE PRODUCT shows only that part of the PRODUCT TYPE hierarchy that contains that term and all narrower terms. The factor term FRUIT JUICE is found at the narrowest hierarchical level.

Any factor term in the CFSAN THESAURUS may be displayed as an individual record. FRUIT JUICE displayed in this way shows the broader term FRUIT JUICE OR RELATED PRODUCT, an alphanumeric code for computer storage, and a definition.

Let's look at another factor. The FOOD SOURCE factor is selected. Some of its broadest categories are ANIMAL USED AS FOOD SOURCE, CHEMICAL FOOD SOURCE, and PLANT USED AS FOOD SOURCE. FRUIT-PRODUCING PLANT, which is hierarchically indented under PLANT USED AS FOOD SOURCE, has a large number of terms indented under it, one

of which is CITRUS FRUIT.

The hierarchical arrangement beginning at the term CITRUS FRUIT is arrayed. Narrower terms include GRAPEFRUIT, LEMON, LIME, and ORANGE; CALIFORNIA VALENCIA ORANGE and SEVILLE ORANGE are indented under ORANGE.

I will now use the CFSAN THESAURUS to select the individual record for ORANGE. The broader term CITRUS FRUIT appears, as well as a Code of Federal Regulations citation and several synonyms, one of which is the scientific name CITRUS SINENSIS. Also displayed are several narrower terms like SEVILLE ORANGE, and some narrower terms like GRAPEFRUIT AND ORANGE that are mixture terms.

Finally, the term ORANGE is an index term for 71 food names. These food names are found in four files. "USDA HANDBOOK 8' is the USDA Nutrient Database for Standard Reference, which was coded with the FFV 2 years ago by the National Cancer Institute in collaboration with CFSAN. This database contains nutrient analytical data. "SIREN" is CFSAN's Scientific Information and Retrieval Exchange Network. This database contains bibliographic citations to FDA regulatory and petition information, some of which may be restricted. CFSAN's "TOTAL DIET STUDY' includes foods that are analyzed for pesticides, toxic elements, and industrial chemicals. "CFSAN THESAURUS food names are built into printed versions of the FFV hierarchy. These names do not retrieve any analytical or bibliographic information.

Factor terms that we've just seen in the CFSAN THESAURUS may now be used to perform searches in the FOOD MONITORING DATABASE.

FOOD MONITORING DATABASE

I will use the FOOD MONITORING DATABASE to search for all food products that are "orange juice". The word ORANGE is entered; 75 food names are retrieved. I add a condition to the search by selecting the term FRUIT JUICE. Retrieval shows 17 SIREN food names, 12 USDA HANDBOOK 8 food names, two TOTAL DIET STUDY food names, and four CFSAN THESAURUS food names.

For the 17 SIREN food names, there are 33 SIREN references; for the 12 USDA food names, there are 487 nutrient/food name combinations; for the two Total Diet Study food names, there are 961 Total Diet residue values. Each database may be searched individually once the Factored Food Vocabulary has provided the link to food products sharing common characteristics.

The SIREN database is selected and three SIREN references are shown. One citation is entitled Sweetener", even though the product that is being indexed is called "Orange juice, frozen concentrated, acid-reduced".

Next, the USDA HANDBOOK 8 database is searched for the nutrient "ascorbic acid'. This information is added to the previous search for all orange juice food names in this file. These search results are displayed in a table listing the food names, number of observations of the food product, the mean value for ascorbic acid, and the standard error.

Finally, the TOTAL DIET STUDY database is searched for the pesticide "carbaryl'. This information is added to the previous search for all orange juice food names. These search results are displayed in a table listing the food names, the pesticide name, the year for which the survey was conducted, and the residue level in parts per million.

In this way, the Factored Food Vocabulary is used to describe foods from a variety of viewpoints. Using the FFV in the Food Monitoring Database permits retrieval of food product names sharing common characteristics and, ultimately, access to nutritional, analytical, or bibliographic data.