

THE FOOD COMPONENT RESEARCH DATA BASE AT THE NATIONAL CANCER INSTITUTE

Pauline M. Herold, M.S., R.D.¹
Ritva R. Butrum, Ph.D.²
Emily M. Brooks, B.A.³
Julio P. Roque, B.S.⁴
Kevin J. Meagher, B.S.⁵
Thomas A. Marciniak, M.D.⁶

INTRODUCTION

Recent advances in nutritional research document that dietary patterns can affect the risk of developing chronic diseases such as cancer and cardiovascular illness (1,2). A diet high in fat, and in particular saturated fat, is a well-established risk factor for cardiovascular disease (1). Diet is postulated to be a factor in the etiology of about a third of all cancers (3).

Information on the nutrient composition of foods is essential in any quantitative study of human nutrition. It is required to assess the nutritional adequacy of diets, to formulate dietary modifications, and to evaluate the role of dietary factors in chronic diseases. Researchers and clinicians need ready access to food composition data of assured quality, accuracy, and appropriateness. The National Cancer Institute's (NCI) Division of Cancer Prevention and Control is developing a personal computer-based data system on food composition that serves as a research tool to its investigators in the field of human nutrition.

The goal of the system, called the Food Component Research Data Base (FCRDB), is to enable the nutrition researcher to easily and efficiently obtain information on the nutritional contents of foods. Compared to other dietary analysis packages, the system in development at the NCI includes standardized descriptions of foods and allows great flexibility with which the user can specify information retrievals. Traditionally, food composition tables have been arranged alphabetically or by food groups. Instead of limiting the user to searches using these classifications schemes, the NCI data system allows users to retrieve information using terms related to multiple characteristics included in the description of each food. The terms comprise a multi-hierarchical faceted food description language called the Factored Food Vocabulary (FFV). (4) Since the use of FFV is integral to FCRDB, it is

¹ Diet and Cancer Branch, National Cancer Institute, Blair Building - Room 619, Bethesda, MD 20892-4200

² Diet and Cancer Branch, National Cancer Institute, Bethesda, Md.

³ Information Management Services, Inc., Silver Spring, MD.

⁴ Intersoft Systems, Inc., Norcross, GA.

⁵ Information Management Services, Inc., Silver Spring, MD

⁶ Computer Systems Section, National Cancer Institute, Bethesda, MD

THE FOOD COMPONENT RESEARCH DATA BASE AT THE NATIONAL CANCER INSTITUTE

printed. CONFIGURE SYSTEM enables users to set up system parameters to match microcomputer, monitor, and printer configuration.

In the SEARCH section, users can search the database for foods with specified characteristics and then produce reports on the nutrient content of these foods. A valid search statement is a Boolean expression where the operand are either FFV factor codes, FFV factor terms, food identification numbers or set numbers, as shown in Figure 2. An on-line FFV dictionary of terms aids users selecting FFV factor terms for a search. The response time is fast: each of the two searches shown in Figure 2 was executed in under four seconds on a standard 6 megahertz IBM PC AT. Foods that have been selected as a result of the most recent search can be viewed, as in the screen which appears as Figure 3.

The REPORT OPTIONS allow users to specify what they want to include in their nutrient content reports. As shown in Figure 4, STATISTICS allows users to select the statistics to appear in the reports. By selecting COMPONENTS, users can make their choice of nutrient and food components to appear in the report. The destination of the report option is selected through OUTPUT OPTIONS. RUN REPORT runs the report.

The seventy-three nutrients and food components in Handbook No. 8 are available in the system. The SELECT COMPONENTS option displays them and allows users to choose amount them. When this option is selected, the first of the two page food component window is displayed. The screen appears as in Figure 5. The components selected can be saved as defaults for subsequent reports.

The system can produce both individual and aggregate nutrient value reports. An individual report provides the statistics for each food in the set individually, while the aggregate report groups the foods in the set together and produces statistics for the group as a whole. The individual report produced for the search requested in Figure 2 appears in Figure 6 and required three seconds to produce. An aggregate report for the same set of foods, taking five seconds to generate, appears in Figure 7. Both types of reports may be printed, displayed on the screen, or sent to a file.

Selection of the DATA MAINTENANCE option of the main menu allows users to enter and edit the FFV descriptions of foods in the data base. The screen for the entry and editing of FFV descriptor codes is shown in Figure 8. The screen displayed as Figure 9 allows users to verify food names and their FFV descriptor terms.

FCRDB

Table 1: FFV Factors and Terms

| <u>Factor</u> | <u>Terms: Approx. # and Examples</u> |
|--|---|
| A) Product Type Manufacturing, consumption and functional characteristics. | 170 dessert bakery product soup |
| B) Food Source Animal, plant or chemical source of food or its major ingredient. | 1005 sugar producing plant shellfish or crustacean |
| C) Part of Plant or Animal Anatomical part of plant or animal from which a food product or its major ingredient is derived. | 120 seed or kernel root, tuber or bulb skeletal meat |
| E) Physical State, Shape or Form Physical state (solid, semisolid, etc.) Characteristics such as viscosity and particle size are also considered. | 45 liquid with low viscosity whole, natural shape |
| F) Degree of Preparation Degree of heat treatment food has had. | 10 uncooked, raw partially cooked |
| G) Cooking Method Cooking method used by the consumer. | 30 cooked by dry heat cooked by microwave |
| H) Treatments Applied (and Ingredients) Processing steps and ingredients in mixed foods. | 180 fat partially removed hydrogenated pickled |
| J) Primary Preservation Method | 45 preserved by fermentation |
| K) Packing Medium Medium in which food is packed for preservation, handling or palatability. | 40 packed in broth packed in water |
| M) Container or Wrapping Main container materials and form. | 100 glass container metal container |
| N) Food Contact Surface Container materials in direct contact with food. | 60 glass plastic |
| P) User Group Age and dietary prescription of the user group for which the food product is marketed. | 20 human food, no age specification, regular diet infants or junior food, regular infant diet |
| Z) Adjunct characteristics The grade, cut, and color of meats. | 35 loin prime grade |

THE FOOD COMPONENT RESEARCH DATA BASE AT THE NATIONAL CANCER INSTITUTE

FCRDB

Table 2: FFV Factoring Examples

Example 1: Skim milk, with vitamin A added

| <u>Code</u> | <u>Description</u> |
|-------------|--|
| A148 | Milk or milk product |
| B1201 | Cow |
| C235 | Milk |
| E123 | Liquid, low viscosity, with no visible particles |
| F18 | Partially cooked |
| G003 | Cooking method not applicable |
| H199 | Fortified |
| H248 | Fat fully removed |
| H213 | Vitamin A added |
| J001 | Preservation method not known |
| K03 | No packing medium used |
| M001 | Container or wrapping not known |
| N01 | Food contact surface not known |
| P24 | Human food, no age specification, regular diet |

Example 2: Canned evaporated whole cow's milk

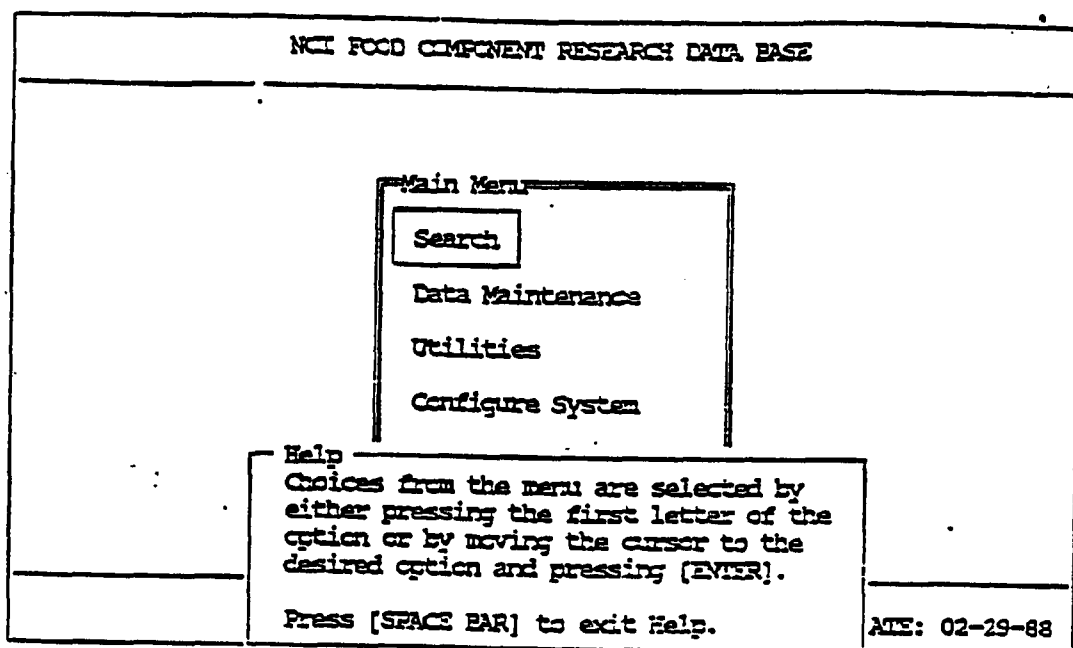
| | |
|-------|---|
| A148 | Milk or milk product |
| B1201 | Cow |
| C235 | Milk |
| E139 | Liquid, high viscosity, with no visible particles |
| F14 | Fully cooked |
| G003 | Cooking method not applicable |
| H114 | Water removed to reconstitution ratio 1 plus 1 |
| J123 | Sterilized by heat, canned |
| K001 | No packing medium |
| M151 | Metal container |
| N24 | Coating enamel |
| P24 | Human food, no age specification, regular diet |

Example 3: Margarine, made from hydrogenated corn oil, sold in a plastic tub

| | |
|-------|--|
| A231 | Margarine |
| B1379 | Field corn |
| C190 | Fat or oil |
| E119 | Semisolid with smooth consistency |
| F14 | Fully cooked |
| G003 | Cooking method not applicable |
| H174 | Hydrogenated |
| H199 | Fortified |
| H213 | Vitamin A added |
| H206 | Alkalized |
| H197 | Bleached |
| J100 | Preserved by adding chemicals |
| K03 | No packing medium used |
| M187 | Plastic container, rigid or semirigid, plastic lid |
| N36 | Plastic |
| P24 | Human food, no age specification, regular diet |

FCRDB

Figure 1: Main Menu



In the SEARCH section, users can search the database for foods with specified characteristics and then produce reports on the nutrient content of these foods. A valid search statement is a Boolean expression where the operands are either FFV factor codes, FFV factor terms, food identification numbers or set numbers, as shown in Figure 2. An on-line FFV dictionary of terms aids users selecting FFV factor terms for a search. The response time is fast: Each of the two searches shown in Figure 2 executed in under four seconds on a standard 6 megahertz IBM PC AT. Foods that have been selected as a result of the most recent search can be viewed, as in the screen which appears as Figure 3.

THE FOOD COMPONENT RESEARCH DATA BASE AT THE NATIONAL CANCER INSTITUTE

FCRDB

Figure 2: Search Statements

| | |
|--|----------------------|
| LAST SET #: 2 | LAST SEARCH HTIS: 12 |
| LAST SEARCH: (a258 and b1312 and j116) and h273 | |
| SET#: 1 ENTER SEARCH STATEMENT: a258 and b1312 and j116 HTIS: 44 | |
| SET#: 2 ENTER SEARCH STATEMENT: set 1 and zinc added HTIS: 12 | |
| SET#: 3 ENTER SEARCH STATEMENT: | |

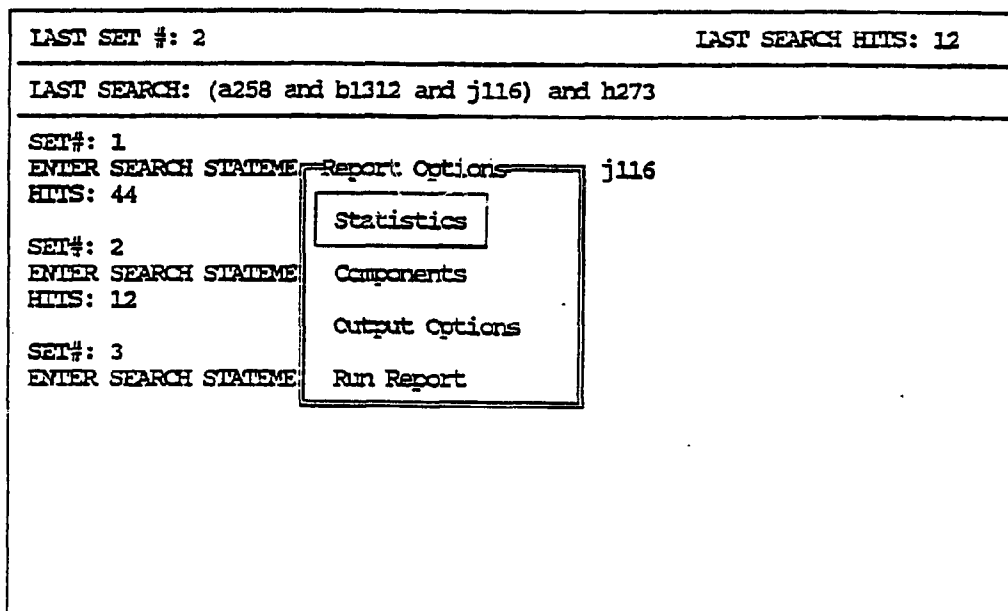
Figure 3: Foods Retrieved by Current Search

| | |
|---|----------------------|
| LAST SET #: 2 | LAST SEARCH HTIS: 12 |
| LAST SEARCH: (a258 and b1312 and j116) and h273 | |
| 08001 ALL-BRAN CEREAL 08005 BRAN BUDS CEREAL 08026 CRISPY WHEATS 'N RAISINS 08028 40% BRAN FLAKES CEREAL, KELLOGG'S 08029 40% BRAN FLAKES CEREAL, POST 08031 FROSTED MINI-WHEATS CEREAL 08036 GRAHAM CRACKOS CEREAL 08051 MOST CEREAL 08060 RAISIN BRAN CEREAL, KELLOGG'S 08061 RAISIN BRAN CEREAL, POST 08073 SUPER SUGAR CRISP CEREAL 08152 WHEAT NUTRI-GRAIN CEREAL | |

FCRDB

The REPORT OPTIONS allow users to specify what they want to include in their nutrient content reports. As shown in Figure 4, STATISTICS allows users to select the statistics to appear in the reports. By selecting COMPONENTS, users can make their choice of nutrient and food components to appear in the report. The destination of the report option is selected through OUTPUT OPTIONS. RUN REPORT runs the report.

Figure 4: Report Options



The seventy-three nutrients and food components in Handbook 8 are available in the system. The SELECT COMPONENTS option displays them and allows users to choose among them. When this option is selected, the first of the two page food component window is displayed. The screen appears as in Figure 5. The components selected can be saved as defaults for subsequent reports.

THE FOOD COMPONENT RESEARCH DATA BASE AT THE NATIONAL CANCER INSTITUTE

FCRDB

Figure 5: Report Component Selection

```

COMPNENTS
-----
PROXIMATE          MINERALS          VITAMINS
Water             Calcium          Ascorbic Acid
Food Energy      Copper           Thiamin
- Protein         Iron             Riboflavin
Total Lipid (fat) Magnesium        Niacin
Carbohydrate     Manganese       Pantothenic acid
Crude Fiber      Phosphoru
Dietary Fiber, Insol. Potassium
Ash              Sodium
Alcohol          - Zinc
Caffeine

Save as new Defaults? (Y/N)

[SPACE BAR] toggles component on/off
[PgDn] for More Components                [ESC] to EXIT
    
```

The system can produce both individual and aggregate nutrient value reports. An individual report provides the statistics for each food in the set individually, while the aggregate report groups the foods in the set together and produces statistics for the group as a whole. The individual report produced for the search requested in Figure 2 appears in Figure 6 and required three seconds to produce. An aggregate report for the same set of foods, taking five seconds to generate, appears in Figure 7. Both types of reports may be printed, displayed on the screen, or sent to a file.

Selection of the DATA MAINTENANCE option of the main menu allows users to enter and edit the FFV descriptions of foods in the data base. The screen for the entry and editing of FFV descriptor codes is shown in Figure 8. The screen displayed as Figure 9 allows users to verify food names and their FFV descriptor terms.

FCRDB

Figure 6: Individual Report

| LAST SET #: 2 | | LAST SEARCH HITS: 12 | | | |
|--|-----------------------------------|----------------------|--------|-----------|-----|
| LAST SEARCH: (a258 and b1312 and j116) and h273 | | | | | |
| Individual Report | | | | | |
| A: Values based on analytical data with no standard error. | | | | | |
| B: Imputed values. | | | | | |
| C: Values based on product label claims. | | | | | |
| USDA-8 ID | Description | # OBS | Mean | Std Error | Msg |
| 08001 | ALL-BRAN CEREAL | | | | |
| | Protein (g) | 3 | 14.300 | . | A |
| | Zinc (mg) | NA | 13.200 | . | C |
| 08005 | BRAN BUDS CEREAL | | | | |
| | Protein (g) | 3 | 13.900 | . | A |
| | Zinc (mg) | NA | 13.200 | . | C |
| 08026 | CRISPY WHEATS 'N RAISINS | | | | |
| | Protein (g) | 6 | 6.900 | 0.111 | |
| | Zinc (mg) | 6 | 1.190 | 0.044 | |
| 08028 | 40% BRAN FLAKES CEREAL, KELLOGG'S | | | | |
| | Protein (g) | 3 | 12.600 | . | A |
| | Zinc (mg) | NA | 13.200 | . | C |

Figure 7: Aggregate Report

| LAST SET #: 2 | | LAST SEARCH HITS: 12 | | | | | | |
|--|-------|----------------------|--------|--------|---------|-----|-----|-----|
| LAST SEARCH: (a258 and b1312 and j116) and h273 | | | | | | | | |
| Aggregate Report | | | | | | | | |
| A: Values based on analytical data with no standard error. | | | | | | | | |
| B: Imputed values. | | | | | | | | |
| C: Values based on product label claims. | | | | | | | | |
| # Miss | # OBS | Min | Max | Mean | Std Err | # A | # B | # C |
| Protein (g) | | | | | | | | |
| 0 | 12 | 6.500 | 14.300 | 10.525 | 2.820 | 9 | 0 | 0 |
| Zinc (mg) | | | | | | | | |
| 0 | 12 | 1.190 | 13.200 | 7.999 | 4.297 | 0 | 0 | 11 |

THE FOOD COMPONENT RESEARCH DATA BASE AT THE NATIONAL CANCER INSTITUTE

Figure 8: FFV Entry Screen

[Pg Up] →. names

Food ID: 01085

Product Type: A148 Food Source: B1201

Part of Plant/Animal: C235 Physical State/Shape/Form: E123

Degree of Preparation: F01 Cooking Method: G003

Treatment Applied: E213 E248 H199 H__ H__ H__ H__ H__ H__

H__ H__ H__ H__ H__ H__ H__ H__ H__

Preservation Method: J001 Packing Medium: K03

Container or Wrapping: M001 Food Contact Surface: N01 N__ N__ N__

User Group: F24 P__ P__ P__ Checklist: Z__ Z__ Z__

Figure 9: FFV Verification Screen

Food ID: 01085 Data Source: USDA _____

Food Name: SKIM MILK, WITH VITAMIN A ADDED _____

Scope Note: RAW AND PASTEURIZED. _____

A148 MILK OR MILK PRODUCT

B1201 COW

C235 MILK

E123 LIQUID, LOW VISCOSITY, WITH NO VISIBLE PARTICLES

F01 DEGREE OF PREPARATION NOT KNOWN

G003 COOKING METHOD NOT APPLICABLE

H213 VITAMIN A ADDED

H248 FAT FULLY REMOVED

H199 FORTIFIED

J001 PRESERVATION METHOD NOT KNOWN

[Pg Dn] → more or [ESC] to quit

DISCUSSION

Functionally, the data system provides a working tool for the nutrition researcher to address complex questions regarding food composition. The interactive implementation on a personal computer with excellent response times makes FCRDB an easily accessible and friendly tool regardless of the level of computer expertise of the investigator. FCRDB also provides more evidence, in addition to the FDA's experience with mainframe systems, that a multihierarchical faceted food composition language such as the FFV is useful in the study of food composition and nutrition research. In fact, the usefulness of complex descriptive languages such as FFV is dependent upon the availability of facile tools such as FCRDB that minimize the complexity from the user's perspective.

Technically, this data system demonstrated that a complicated descriptive language can be used for complex retrievals on a personal computer with excellent response times. FCRDB applies current, state-of-the-art software techniques (C, windowing support, B-tree handlers) with appropriate programming and data structures (e.g., bitmaps) to achieve excellent performance on a small machine. Previous applications of the FFV require the use of mainframe computers. FCRDB is an example of optimizing a medical application for the personal computer environment by applying state-of-the-art, yet standard software techniques.

Both functionally and technically, FCRDB has a rich future. The FFV is being proposed for international standardization. The future standardization efforts will make FFV more relevant to a wider range of users and eliminate some of the current limitations in its application. Technically, the initial implementation of FCRDB included code entry because the initial emphasis was upon demonstrating that the retrieval performance on a personal computer using the FFV would be adequate. Now that that question has been eliminated, the user interface is being optimized to provide easier use for researchers not intimately familiar with FFV codes. We view the NCI data base as it exists today as a small module on which to build future expansions. Integrating it to other personal computer based food and nutrition systems at the NCI presents an exciting functional and technical challenge.

REFERENCES

1. Pyorala, K. Dietary cholesterol in relation to plasma cholesterol and coronary heart disease. Amer Jour Clin Nutr 1987 45:1176-1184.
2. Palmer, S. Diet, nutrition and cancer. Progress in Food and Nutrition Science 1985 9:283-341.
3. Doll, R., Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. Jour Nat Cancer Inst 1981 66 (6): 1191-1308.
4. McCann, A., Pennington, J.A.T., Smith, E.C., Holden, J.M., Soergel, D., Wiley, R.C. FDA's Factored Food Vocabulary for food product description. Jour Amer Diet Assoc 1988 88(3) : 336-341.
5. Harris, L.E., Haendler, H., Riviere, R., 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. Publication 2. Prepared on behalf of INFIC by the International Feedstuffs Institute, Utah Agricultural Experiment Station, Utah State University, Logan, Utah, U.S.A., 1980