

## DOCUMENTATION OF DATA SOURCES

Patricia D. Catsos

*Tufts Nutrient Data Center*

On behalf of the Nutrient Database Committee, I am here to lead a discussion on the documentation of data sources for nutrient databases. Developers of some databases use systematic methods of identifying sources of nutrient data, coding rules, recipes, portion weights, imputation methods and other database characteristics. The documentation system may be built into the database software, may take on the form of supporting paper documents, or both. A data documenting system is used to assure continuity despite staff turnover, to describe the database to users, and to assist with updating. The need for such a system becomes more pronounced with a large or long-lived database.

The Tufts Nutrient Database uses "reference codes" in the database itself to indicate sources for:

- a) The primary source of nutrient data for a food.
- b) Nutrient source, if different from the primary source.
- c) Portion weight source, if different from the primary source.

When large batches of data are read into the database (i.e., USDA tapes), computer programs assign the reference codes as appropriate. Reference codes can also be added or changed individually, using the database editing program.

At the heart of our data documenting system is a 160-entry bibliography, which is maintained with the data bank management system INGRES. This method is efficient because each source is described in detail only once. Each bibliography entry corresponds to one of the reference codes which is used within the database. To combine references or describe imputation methods, a memo is written and assigned a reference code of its own. (See attached printout of the Tufts Nutrient Database bibliography.)

A library is maintained which corresponds to the bibliography. All books, journal articles, photocopies, worksheets, food labels or government publications are numbered and are available for review. This is useful for updating data and for verifying data accuracy.

In the spring of 1988, a survey was distributed to all contributors of the 1987 Nutrient Databank Directory. Twenty-nine responses were received, which represented 38% of the questionnaires sent out (see attached list of respondents). The survey results are included here for your review.

DOCUMENTATION OF DATA SOURCES

Number	Author(s)	Description	Source	Date
52	Nutrition Coordinating Unit	Frequency of Food Codes in Diet Records	Unpublished	1906
53	Leung W, Butrum RR, Chang HJ, Rao HH, Poljacchi V	Food Composition Table for Use in East Asia	USDA	1972
54	Veltrauch J	Food Consumption: Households in US...HCS 1977-78 No. II 6	USDA	1978
55	Exler J, Veltrauch J	Prev Table on Fatty Acid and Chol Content of Selected Foods	USDA	1978
56	Paul AA, Southgate DAT, Russell J	Prev Table on Content of Omega-3 FA and other Fat Components...	USDA	1978
57	Heplum FH, Exler J, Veltrauch J	First Suppl to McCann and Widdows's The Composition of Foods	Elsevier/NL Holland Biomedical Press, NY	1980
58	Foxall LP, Finsella JA, Watt BK	Fats in the Diet: Why and Where	Food Technology October Issue	1980
59	Anderson DA, Kinsella JA, Watt BK	I. Dairy products. Comp eval of fatty acids in foods	JADA 66:687-696	1975
60	Fristrom CA, Stewart BG, Veltrauch H, Foxall LP	II. Beef products. Comp eval of fatty acids in foods	JADA 67:351-355	1975
61	Exler J, Finsella JA, Watt BK	III. Nuts, peanuts and soups. Comp eval of fatty acids in foods	JADA 68:226-229	1976
62	Exler J, Finsella JA, Watt BK	IV. Unhydrog fats and oils. Comp eval of fatty acids in foods	JADA 68:315-360	1976
63	Exler J, Finsella JA, Watt BK	V. Cereal products. Comp eval of fatty acids in foods	JADA 69:263-268	1976
64	Exler J, Finsella JA, Watt BK	VI. Fish. Comp eval of fatty acids in foods	JADA 69:517-522	1976
65	Exler J, Finsella JA, Watt BK	VII. Eggs. Comp eval of fatty acids in foods	JADA 70:53-58	1977
66	Exler J, Finsella JA, Watt BK	VIII. Lamb and veal. Comp eval of fatty acids in foods	JADA 71:68-72	1977
67	Exler J, Finsella JA, Watt BK	IX. Sausages and luncheon meats	JADA 72:322-326	1977
68	Exler J, Finsella JA, Watt BK	X. Fat, Chol, and Prev Anal of Some Ready to Eat Foods	JADA 66:160-166	1975
69	Exler J, Finsella JA, Watt BK	XI. Major fatty acids and proximate comp of dairy products	JADA 66:11-13	1975
70	Exler J, Finsella JA, Watt BK	XII. Lipids and FA in Important Fishes New Data for Nutrition Tables	JADA 66:11-13	1975
71	Exler J, Finsella JA, Watt BK	XIII. Provisional Table on the Nutrient Content of Fast Foods	JADA 66:11-13	1975
72	Exler J, Finsella JA, Watt BK	Prev Table on the Nutrient Content of Bakery Foods...	USDA	1981
73	Exler J, Finsella JA, Watt BK	Nutrient Content of Beverages	USDA SEA	1981
74	Exler J, Finsella JA, Watt BK	Folate in selected foods	USDA	1981
75	Exler J, Finsella JA, Watt BK	Prev Table on Percent Retention of Nutrients in Food Preparation	JADA 70(7):161-177	1977
76	Exler J, Finsella JA, Watt BK	Lipids in Margarine and Margarine-Like Foods	USDA	1981
77	Exler J, Finsella JA, Watt BK	General Foods	USDA	1981
78	Exler J, Finsella JA, Watt BK	Foody Rich Cons Cent, Box 7100, Madison, WI	USDA	1981
79	Exler J, Finsella JA, Watt BK	Carl Publ'g, 1914 So. Peoria St, Chicago	USDA	1981
80	Exler J, Finsella JA, Watt BK	Samline Bldg, Box 370 Hinn, Hinn	USDA	1981
81	Exler J, Finsella JA, Watt BK	Hershey, Box 805 Hershey, Pa	USDA	1981
82	Exler J, Finsella JA, Watt BK	Red Bull Info Ctr, Oak Brook, Ill	USDA	1981
83	Exler J, Finsella JA, Watt BK	JB Lippencott Co, New York	USDA	1981
84	Exler J, Finsella JA, Watt BK	Deputy Books, Huntington, NY	USDA	1981
85	Exler J, Finsella JA, Watt BK	University of Massachusetts, Amherst, MA	USDA	1981
86	Exler J, Finsella JA, Watt BK	American Egg Board, Park Ridge, Ill	USDA	1981
87	Exler J, Finsella JA, Watt BK	Local	USDA	1981
88	Exler J, Finsella JA, Watt BK	Local	USDA	1981
89	Exler J, Finsella JA, Watt BK	Local	USDA	1981
90	Exler J, Finsella JA, Watt BK	Local	USDA	1981
91	Exler J, Finsella JA, Watt BK	Local	USDA	1981
92	Exler J, Finsella JA, Watt BK	Local	USDA	1981
93	Exler J, Finsella JA, Watt BK	Local	USDA	1981
94	Exler J, Finsella JA, Watt BK	Local	USDA	1981
95	Exler J, Finsella JA, Watt BK	Local	USDA	1981
96	Exler J, Finsella JA, Watt BK	Local	USDA	1981
97	Exler J, Finsella JA, Watt BK	Local	USDA	1981
98	Exler J, Finsella JA, Watt BK	Local	USDA	1981
99	Exler J, Finsella JA, Watt BK	Local	USDA	1981
100	Exler J, Finsella JA, Watt BK	Local	USDA	1981
101	Exler J, Finsella JA, Watt BK	Local	USDA	1981
102	Exler J, Finsella JA, Watt BK	Local	USDA	1981
103	Exler J, Finsella JA, Watt BK	Local	USDA	1981
104	Exler J, Finsella JA, Watt BK	Local	USDA	1981
105	Exler J, Finsella JA, Watt BK	Local	USDA	1981
106	Exler J, Finsella JA, Watt BK	Local	USDA	1981
107	Exler J, Finsella JA, Watt BK	Local	USDA	1981
108	Exler J, Finsella JA, Watt BK	Local	USDA	1981
109	Exler J, Finsella JA, Watt BK	Local	USDA	1981
110	Exler J, Finsella JA, Watt BK	Local	USDA	1981
111	Exler J, Finsella JA, Watt BK	Local	USDA	1981
112	Exler J, Finsella JA, Watt BK	Local	USDA	1981
113	Exler J, Finsella JA, Watt BK	Local	USDA	1981
114	Exler J, Finsella JA, Watt BK	Local	USDA	1981
115	Exler J, Finsella JA, Watt BK	Local	USDA	1981

116	Augustine	Augustine International, Ltd, Balaoy, NJ	1987 HRC
117	Estee, Dia Hel	Estee Corporation, Parsippany, NJ	1987 HRC, HRC
118	Felischman's, Hablisco	Hablisco Consumer Info, East Hanover, NJ	1987 HRC, HRC
119	Underwood, Accent	Pat Incorporated, St. Louis, MO	1987 HRC, HRC
120	Peabier	Peabier Company, Elmhurst, IL	1987 HRC
121	King Oscar	King Oscar Inc, Westport, CT	1987 HRC
122	AC	AC Foods Corp, Brooklyn, NY 11208	1987 HRC, FC
123	Mrs. Bach, Mrs. Bach Hotline	Alberto Culver Co, Reticon Park, IL	1987 HRC
124	Stop and Shop	Stop and Shop Co, Boston MA	1987 HRC, HRC
125	Weight Watchers	Weight Watchers Int., Hahnbassel, NY	1987 HRC
126	Monte	Monte Ice Company, Inc, Fullerton, CA	1987 HRC
127	Lorraine	Lorraine Cheese Co, Pittsfield, IL	1987 HRC
128	Saffola	Saffola Co, Los Angeles, CA	1987 HRC
129	Chiffon	Chiffon Foods, Div of Anderson Clayton	1987 HRC
130	Reisman	J. Reisman & Sons Co, Inc, Pennsauken, NJ	1987 HRC
131	McGladwin	McGladwin Cheese	1987 HRC
132	Veemanen RI, Sedor FA, Kimm STS	JAVA 02:374-396	1987 HRC
133	Featherlight	Cable Diet Supply Inc, LaGrange, IL 60525	1987 HRC
134	Creus, Barbara	DM, Fronghe Inc, New York, NY	1987 HRC
135	Complete Guide to Sodium	Stella Doro Biscuit Co, Inc, Bronx, NY	1987 HRC, HRC
136	Stella D'oro	Stella D'oro Biscuit Co, Buffalo, NY	1987 HRC
137	San Francisco Heart Association	Besser Foods, Buffalo, NY	1987 HRC
138	DeJoseph	San Francisco Heart Assoc, Inc, SF, CA	1987 HRC
139	Alka Seltzer	Alkette Balmich & Co, Fullerton, NJ	1987 HRC
140	Brown Seltzer	Hiles Laboratories, Elkhart, IN	1987 HRC
141	Sweet & Low	Warner Lambert Co, North Plains, NJ	1987 HRC
142	Heinz	Heinz Food Industries, Inc, Evanston, IL	1987 HRC, HRC
143	Riverside Methodist Hospital	Heinz Company, Pittsburgh, PA	1987 HRC
144	Rehseby MR, Gatto AH, Scott LV, Froynt JB	River-De Both Hosp, Columbus, OH 43214	1987 HRC
145	Imputed by computer program. Caffeine presumed-0.	Simon & Schuster, NY	1987 HRC
146	Imputed by computer program. Alcohol presumed-0.	Coca-Cola	1987 HRC
147	Classic Lite	Coca-Cola Co, Atlanta, GA	1987 HRC, HRC
148	Camp of Foods, FinFish and Shellfish Products, All 0 15	Almond Foods	1987 HRC, HRC
149	Item to Rvgn Under Omega 3 Fatty Acids	USDA HHS	1987 HRC, HRC
150	Edler J		1987 HRC
151	Catsos, Fatsy		1987 HRC
152			1987 HRC

**P.D. CATSOS**

**SURVEY RESPONDENTS**

1. Albert Einstein College of Medicine
2. American Health Foundation
3. Case Western Reserve University
4. CBORD
5. Colorado State University
6. Computrition
7. DFM Software Systems
8. Dietary Data Analysis
9. ESHA Research
10. Godin London, Inc.
11. Golden West College
12. Massachusetts Nutrient Database
13. Medical College of Wisconsin
14. Micromedx
15. N-Squared Computing
16. Nutriquest II
17. Nutrition and Diet Services
18. Nutrition Computer and Statistical Service
19. Nutrition Coordinating Center
20. Nutrition Scientific
21. Nutritional Data Resources
22. Ohio State University Hospitals
23. Quilchena Consulting Ltd.
24. Softech Computing Company
25. Syracuse University
26. UCLA
27. University of California, Berkeley MiniList
28. University of Nevada
29. University of Pittsburgh

## DOCUMENTATION OF DATA SOURCES

### SURVEY RESULTS

1. Can the following sources of data be identified for each data bank entry? (Check all that apply).
  - 27 Primary source of data for food item
  - 20 Individual nutrients if different from the primary source
  - 18 Portion amounts/weights if different from the primary sourceOther:
  - Comments/free text regarding averaging, source selection
  - Estimated values indicated
  - Difficult to "define" combinations of data sources
  
2. Are the data sources referenced within the data bank itself or on paper (Check one).
  - 18 Within data bank
  - 1 Other computer-readable file
  - 18 On paperComments:
  - Combination frequently used
  - Paper system has slow access, but is simple
  
3. If "Within data bank," are paper files maintained which correspond to the references within the data bank? (Check one).
  - 14 Yes
  - 2 NoComments:
  - Also computer readable tapes, e.g. USDA Database for Standard Reference
  
4. How are data sources referred to? Check one)
  - 10 Information contained in the food code itself
  - 17 Bibliographic reference number used
  - 10 Other:
    - Category code stored with nutrient data
    - Source sheet/documentation page for each food item
    - In user's guide/log book
    - Estimated values are underlined
    - Descriptive fields
    - Alphabetical listing with cross-reference
  
5. How many data sources are referred to in the current version of your database? (Check one).

9	0-10
10	11-50
4	51-100
2	101-200
2	> 200

Comments:
  - more than 400
  
6. How specific are your references? (Check all that apply)

**P.D. CATSOS**

- 22 Type of data source specified (i.e, label claim vs. journal vs. USDA tape)
- 13 Rationale for imputed data indicated
- 19 Company names and addresses for manufactured foods
- 19 Book titles, authors, publishers
- 17 Journal article titles, authors, page numbers
- 2 Other:
  - Actual vs. imputed vs. user input value indicated
  - Each imputed value has its own code
  
- 7. Applications for your referencing system: (Check all that apply)
  - 24 Updating nutrient and non-nutrient data
  - 24 Verifying data accuracy
  - 22 Answering client inquiries about sources of data
  - 7 Selecting which foods, nutrients to use in a given analysis
  - 3 Other:
    - Addition of non-USDA foods to database
    - List references in publications that contain data analyzed by system (11-20 refs.)
    - Determining % imputed values
    - Determining % values from a source
  
- 8. Are there other relevant facts about your referencing system that you would like to share?
  - No client inquiries about sources in past 6 years
  - Referencing system created as an"appendage" for another system