

# **USDA Nutrient Databank**

David Haytowitz, USDA-ARS-HNIS

## **Introduction**

There is an ancient curse which states "May you live in interesting times." The past year, with all the changes it has brought, has certainly been interesting. With change comes opportunities. Once a separate agency within the USDA, the Human Nutrition Information Service (HNIS) is now part of the Agricultural Research Service (ARS). We have worked extensively with our colleagues in ARS in the past and look forward to working even more closely with them in the future.

In addition to being moved in the departmental organization chart, we are also physically moving. Our new offices will be in Riverdale, which is a couple of miles from our current location. By next year we will know our new address and phone numbers. In the meantime we will keep you informed by posting messages on the Internet and other bulletin boards.

Last October, the General Accounting Office's (GAO) report entitled "Better Guidance Needed to Improve Reliability of USDA's Food Composition Data" was released. In another paper in these proceedings, Betty Perloff describes GAO's recommendations and the steps we are taking to implement them.

This paper covers some of the activities of the Nutrient Data Research Branch, which is responsible for compiling representative food composition values. These values are made available to researchers around

the world both in printed form as Agriculture Handbook No. 8 (AH-8) and its machine-readable form, the USDA Nutrient Data Base for Standard Reference. The data are also used in specialized data sets for the food surveys conducted by USDA, the NHANES survey conducted by the DHHS, and the National Nutrient Data Base for Child Nutrition Programs.

### **Bulletin Board**

The Nutrient Data Bank Bulletin Board, which has been operating since 1990, continues to serve as a method of data dissemination and announcements about HNIS publications and the Nutrient Data Bank Conference. It is accessible by individuals and institutions and enables them to transfer nutrient data directly to their own computers. This information is updated and revised monthly.

The board operates 24 hours a day, 7 days a week, and averages over 350 callers per month from nearly every state in this country and some foreign countries including Australia, Japan, and others. In addition to the files described below in greater detail, some of the other files currently on the Bulletin Board include data from Home and Garden Bulletin No. 72, "Nutritive Value of Foods, Home Economics Research Report No. 48, "Sugar Content of Selected Foods", and the Provisional Tables on Selenium and Vitamin D. There are also data sets containing bibliographic and other information of interest to nutrient data users prepared by the Food and Nutrition Information Center at the National Agricultural Library. The telephone number for the Bulletin Board is 301-436-5078.

### **Internet**

In cooperation with the University of Maryland, the data and information on the bulletin board are also available over the Internet. To get to our data type:

telnet info.umd.edu

at your system prompt and select the following items from the menu:

- Educational Resources
- Government
- United States
- Executive
- Nutrient Data

Some of you may have encountered some problems finding our data recently. This happened when the University of Maryland reorganized their directories. The new location of our data in the directory structure, which is only slightly different from before is listed above. If, in the future, you have problems finding our data, send me a message on the Internet at:

info-12@info.umd.edu

### **Progress**

#### **Release 10**

Release 10 of the USDA Nutrient Data Base for Standard Reference was released in July 1993 and made available at NTIS, on the Nutrient Data Bank Bulletin Board, and on the Internet. This release added new

data from AH-8 sections on Baked Products (AH-8-18), Snacks and Sweets (AH-8-19); Fresh Pork, and the 1990 and 1991 supplements. In addition, reflecting improvements in the availability of data, crude fiber data was dropped and replaced with data on total dietary fiber (TDF). While TDF data was available for only about 3000 items, this information will be updated and expanded in future releases.

### **AH-8 Supplements**

HNIS continued updating "Composition of Foods...Raw, Processed, Prepared," Agriculture Handbook No. 8 (AH-8) by issuing last year the fourth in a series of supplements to the handbook. These supplements give us an opportunity to add new data which have resulted from contracts and data collected from other sources as well as updating items previously published in the 21 sections of AH-8. This last supplement contains new or revised data for 59 food items in 11 sections of AH-8. These data will be incorporated into the next release of the USDA Nutrient Data Base for Standard Reference. For those users who wish to access the data now, it is available on the Bulletin Board. The 1993 supplement is in preparation and will contain data on cooked beef and lamb cuts trimmed to 1/8" external fat. Planning for the 1994 supplement has begun. It will include data on processed eggs and data on other foods from current contracts.

### **New Beef and Lamb Data**

In February 1994, data on 56 cooked beef and 14 cooked lamb cuts trimmed to 1/8" external fat was made available on the electronic bulletin board. These data on a 100 gram basis only were released electronically so they could be used for nutrition labeling. Based on market studies the Food Safety and Inspection Service (FSIS) of the USDA is requiring that the cooked edible portion consisting of lean and fatty tissue of meat cuts trimmed to 1/8" fat be used for nutritional labelling. However, AH-8 contained meat data on 1/4" and/or 0" trim. Therefore a team of cooperators was organized by HNIS and the National Live Stock and Meat Board to produce nutrient values for 1/8" trim external fat for beef and lamb cuts. The cooperators were HNIS staff, meat scientists from Texas A&M University, a statistician from the University of Maryland, and representatives from the National Live Stock and Meat Board and FSIS.

Based on fat trim data collected from a market basket study, regression equations were developed for raw meat cuts which predicted the amount of fatty tissue lost due to trimming from 1/4" trim external fat to 1/8" fat. The nutrient composition of 1/8" trim raw meat cuts were then calculated from those equations, and cooked composition data were developed from the raw data. The meat industry will use these data to implement its point-of-purchase nutrition labeling program "Nutri-Facts<sup>TM</sup>," as required by the USDA.

### **Survey Nutrient Data Base**

The USDA Nutrient Data Bases for Individual Food Intake Surveys used in the 1989, 1990, and 1991 Continuing Survey of Food Intakes by Individuals were also released. These data bases were also used for analysis of NHANES 3, phase 1, data. The Primary Nutrient Data Set (PDS), the Nutrient Retention Factors, and the Recipe File, which contain data used to create each version of the survey nutrient data base were released at the same time. Beginning with the next release (CSFII 1994) the formats for the data base will change. Lois Steinfeldt discusses the new formats and some helpful information on using them elsewhere in these proceedings.

## National Nutrient Data Base for Child Nutrition Programs (NNDCNP), Release 1

A new activity which we have undertaken in cooperation with the Food and Nutrition Service is development of the National Nutrient Data Base for Child Nutrition Programs. It will be used in a demonstration project of Nutrient Standard Menu Planning (NSMP) starting in the 1994-95 school year. The NSMP will require that school meals meet the goal for selected nutrients of 1/3 of the RDA for lunch or 1/4 of the RDA for breakfast as well as not more than 30% of total calories from fat. In the NSMP demonstration project, participating schools will analyze the nutritional composition of breakfast and lunch menus utilizing a software package which incorporates the NNDCNP. Software packages are being prepared by a variety of vendors. The NNDCNP was described in greater detail at a workshop held for potential vendors and others during last year's conference in Baton Rouge.

The data base will contain data on 15 nutrients from the Standard Reference data set as well as brand name data submitted by food processors. To facilitate the submission of processed food data, an electronic form was developed which has been sent to several hundred food processors. This form enables the food processor to enter the data on a diskette and submit it to HNIS for inclusion in the data base. For companies supplying large amounts of data, information is available on how they can export the data from their own data bases and send it to us in ASCII format. Along with the nutrient data, companies are also required to submit information on the quality control procedures used to insure the accuracy of the data. Once the data are submitted we will check it for completeness and accuracy. These data will become available in subsequent releases of the data base.

Release 1 of the NNDCNP data files was made available on the Bulletin Board earlier this year. Accompanying the files was a document describing file formats.

### Vitamin K

A revised Provisional Table on the Vitamin K Content of Selected Foods was also released, and copies are available from our office. The revised table contains vitamin K values for 194 food items analyzed by HPLC. These values replace and add to those published in 1990. Values analyzed by bioassay in the 1990 table have been dropped. A corresponding data set is also available on the Bulletin Board.

### Contracts

We have a couple of important contracts under way. One is analysis of the proximate, mineral, and vitamin content of selected ethnic foods such as Puerto Rican cheeses, several peppers, rice paper, and taro stems. A companion contract covers fatty acids and sterols for the same foods. We have received data from the first year of these contracts and work is well under way on the second year. Depending on the availability of funding, we may have a third year and will be able to analyze additional foods. These data are very helpful in assessing the diets of survey respondents for CSFII and NHANES from various special population groups. The data will be released as part of the survey nutrient data sets and in future releases of the Standard Reference Data Base.

Another contract, which will be completed shortly, covers dietary fiber and individual sugars for 85 high-consumption foods. Total dietary fiber will be determined by summing the values for soluble and insoluble fiber. Sugars being analyzed individually by HPLC include glucose, fructose, sucrose, maltose, and lactose. The fiber data will be used to update the primary data set and the Standard Reference Data Base. The sugar data will be available in publications and on the Bulletin Board.

## Plans

### Nutrient Data Bank System Redesign

Another major activity is the redesign of the Nutrient Data Bank System. Since we started using the current system in 1985, many technological advances have occurred and we have undertaken additional activities which cannot be adequately supported by the current system. This redesign will give us a new state-of-the-art system utilizing a relational database management approach, enabling us to more efficiently meet our current data processing needs and providing the flexibility to add new functions when required.

Dr. Loretta Hoover spent 6 months with us last year reviewing the current system, suggesting areas for improvement, and working with staff to develop design objectives for the new system. A users' group comprising other nutrient data users in the Federal Government was also formed last year, and has met twice. During one of its meetings Dr. Hoover presented some of the results of her work with us.

A number of teams were formed to work with Dr. Hoover on a variety of subjects relating to the NDBS redesign. After discussing a wide range of issues relating to these subjects, a number of working papers were developed in the following areas: analytical methods, calorie factors, commercial products (food labels) support file, data screening and quality evaluation, data derivation codes, fatty acids and lipid factors, formulations, identification numbers and data fields, nitrogen-to-protein conversion factors and protein quality, data integrity equations, and retention factor files. In addition recommendations for statistical requirements for the new system were provided by the National Agricultural Statistics Service. Meetings were also held to discuss quality issues. These working papers and recommendations will guide us in the redesign of the system.

As part of her assignment Dr. Hoover looked at how other countries were using computer-based systems for managing their food composition tables. During her investigations she discovered EuroNIMS, a data management system for food composition data developed for six European countries. Dr. Thierry Arnouts from Belgium was invited to present the EuroNIMS system to the NDRB staff. During his visit discussions were held to determine how this system could meet USDA and user needs. Hardware and software to support the EuroNIMS system has been ordered. This will enable us to conduct "hands on" testing, as well as determining if EuroNIMS will meet our needs. During the testing phase we will also identify where enhancements may be needed to meet the requirements identified in the working papers described earlier. Adoption of the EuroNIMS software will result in considerable cost savings, not to mention savings in the time needed to design, write, test, debug, and install a new system. It will also facilitate the exchange of data with other countries using the system.

We also visited FDA for a demonstration of the Languag coding system, the International Interface Standard for Food Databases, and software being developed under contract for FDA by TAS, Inc. Data fields in the Food Labeling and Packaging Survey (FLAPS) conducted by FDA were also compared with the Commercial Products Support File which we will incorporate into the revised system.

This system redesign will enhance NDRB's continuing efforts in the development of representative food composition values for users within the Government and outside. The redesign will also take advantage of improvements in computer hardware and software technology.

## Release 11

Release 11 of the USDA Nutrient Data Base for Standard Reference is planned for release in late 1994. It will include data from the 1992 and 1993 Handbook No. 8 Supplements. Data on the vitamin E content of foods, derived from various forms of the vitamin, will also be included and reported as milligrams alpha-tocopherol equivalents.

### Format

As well as adding new data, we will adopt a new format for the Standard Reference data set. The new format will allow easier importing into a variety of applications such as data base management systems, spreadsheets, statistics programs, and others. This new format will be very similar to those presented last year in Baton Rouge for the Survey Nutrient Data Sets and the Nutrient Data Base for Child Nutrition Programs. However, there will be additional fields in order to present some information unique to the Standard Reference Data Base. These include a new field for source information which is now contained in the standard error field. In the new format, the standard error field will only contain the standard error. A new field will be added designating the food group. In the future, this format will allow us to provide more detailed information requested by our users. Initially, some of these fields may be blank or contain default values, but as this improved data base matures, more information will become available.

This new format will also include more complete food descriptions. This will eliminate the need for abbreviations in the long names. The short names will also have fewer, easier to understand abbreviations. There will be the capability for additional household descriptions and weights. Initially, there may only be the existing two household measures, but more will be added in subsequent releases. Plans are underway to add another file which would include the text of footnotes which appear on many of the printed pages of Agriculture Handbook No. 8. Much of these data were previously unavailable to users of the machine-readable data sets. Once the format is finalized, we will provide documentation along with a sample data set on the bulletin board.

### Interim Releases

In order to disseminate data to you and our other users in a more timely manner, we will begin issuing interim releases to the standard reference data base. This will enable us to make minor updates and corrections without waiting for one of the supplements or a major release. The exact timing of the interim releases will depend on the amount of data available and staff resources. The interim releases will be numbered 11.1, 11.2, and so forth. The interim releases will only be available on the bulletin board and Internet. With each interim release, we will provide the complete standard reference data set with the updates and additions incorporated. A second data set will contain only those foods items which changed with the interim release. Only major releases (11.0, 12.0, ...), occurring approximately once a year and coinciding with the publication of a supplement, will be sent to the National Technical Information Service (NTIS) for sale. Once a major release is made available, the previous interim releases will be removed.

### Computer Programs

A new in-house program has been written to improve the ability of our staff to query the USDA Nutrient Data Base for Standard Reference, and we are looking for a few beta testers to help us evaluate the program. Queries can be made by food description or by NDB Number. The queries can be limited to specific nutrients or food groups. Output from the queries can be displayed on the screen or saved in a file. This file can then be printed or imported into another application. If you are interested in becoming a beta

tester for this program, let me know and we will send you a copy. This program uses the current format of Standard Reference.

As always, we welcome your suggestions as to how we can improve any of the products and areas I have discussed and look forward to your input.