

Management of Food Composition Data Bases

Brucy C. Gray
HNIS / USDA
Hyattsville, Maryland

The first item of consideration in the management of a database is the use to which the database is to be put. A database in a retail store may be used to obtain the price of an item as it is sold and to maintain an inventory of the stock on hand. Except for the stock on hand and price, the actual values of fields in the database remain unchanged throughout its existence. Any new item in that product line would be inserted as a new item in the inventory although it may be a slight modification of an existing item. In this usage, the user system needs the capability of reading all fields in the file and updating the inventory count field and the cost field. The user system, along with its other tasks, needs an efficient way of adding new items to the file and locating a particular record in the file.

Another type of data base is one which is quite dynamic, needing to be updated on a continuous basis. Such a database may be used by a researcher or a medical doctor who maintains an automated record of his patients and the history of an ailment along with the effect of a treatment that is being provided. Such a database as this needs the capability of being constantly updated with expanded field sizes and the capability of comparing entries for different records for the same individual along with making comparisons across individuals. The need for instant access to a particular record is not as great as that needed for the retail store sales but is greater than what is provided by sequential reading. For the convenience of making comparisons, it may be necessary to translate the data being entered into predefined special codes. There is a need to be able to enter text fields as well as a capability of pulling up a specific text field for review at the request of the user. There may be a need to retrieve all records from the data base for the cases where predefined fields contain specific codes or responses.

The message in the above two examples is that there is no single database or response that meets all of the needs of the users. Each application needs to be analyzed and a system developed which best meets those needs. This also implies that the management of the database is correlated with the way the database is set up and the uses to be made of that database. The remaining portion of this presentation will deal with the specific applications of the food composition databases that we have in HNIS.

The most popular of the databases that we distribute is the USDA Nutrient Data Base for Standard Reference. This is the machine readable form of the data contained in the Agriculture Handbook Number 8 series. The database itself is created from files produced in the Agency's Nutrient Data Bank system. That system has been discussed many times and I will not go into details about it at this time. Yesterday afternoon Betty Perloff gave an update to the status of that data bank.

Out of that databank and the system for producing the distributed copy, we receive the necessary files for the data and for the coding manuals. An automated system exists for producing files in the format for distribution. This is a rather straight forward system requiring only that the master file be retained in a machine readable form. The variable form of the system resides in the databank itself.

Two other composition data bases that we distribute for public use are the USDA Nutrient Data Base for Individual Food Intake Surveys and the USDA Nutrient Data Base for Household Food Use Surveys. The system of maintenance for these two are quite similar. The procedure for creating and maintaining the machine readable form of these data bases is automated. Its operational principle is based on the ability to determine the nutrient content of foods by starting with

the nutrient content of the raw ingredients and obtaining the nutrient content of the finished product by appropriate adjustments to account for the gains and losses that occur in the preparation process. The automated form of this system consists of files of all foods that may be used as ingredients in recipes, and the developed automated procedure for processing the files.

At the start of a survey, files exist that contained all of the foods anticipated to be consumed by participants in the survey. For all raw ingredients, the file contains the nutrient content of the ingredient. For foods that are not consumed as raw ingredients a recipe file exists which contains the ingredients used in the prepared food, the respective quantity used, the appropriate refuse factor, the nutrient retention code and other information that impacts on the final nutrient level in the consumed food. The system is structured around performing the needed computations and producing a finished record of the nutrient content of the eatable product. To use this recipe file a system has been designed which takes the ingredients in each recipe and calculates the nutrients in the finished product after adjusting for moisture gain or loss, nutrient loss due to heat and the absorption or loss of fat or other substances.

In this system, the format of the files is fixed. There is a need for immediate access to the records of the ingredients, the supporting files of the nutrient retentions and the refuse factors. There is also a requirement to access the files of the system as needed. This system must also have the capability of being updated as needed with the capability of using the result of previously generated recipes and ingredients of new recipes. This system differs from the two mentioned above in the sense that a part of its structure is a well defined computational algorithm. It takes in information in a fixed format and, as needed, puts out information in a fixed format. The system does not have the capability of expanding the field sizes in order to hold more information and the retrieval of information is not based on the description of the food.

In summary, databases differ based on their content, their size, as well as the intended use. Because of the sizes of the files and the specialized computations that are required, ours have been developed to permit easy handling. It permits them to be stored in an inexpensive form and gives the desired control for maintenance. This has been done by writing specialized programs to perform the handling. Applications centered around the use of Database Management Packages can usually do all that is needed. The

drawback on that is that the applications systems have to be developed and the cost of maintenance is greater than the cost of maintaining sequential files. The guidance that I would give is to discuss the files and the manner in which they will be used with a person experienced in computer operations and let that person provide guidance on a system suited for the applications intended.