INFOODS UPDATE: 1987

THE GOALS OF INFOODS

What INFOODS is attempting to do is to add an option to:

* create an environment for *horizontal* consistency within each of the four groups, so that information (data) can be shared and, at the very least, workers can tell each other, easily and unambiguously, what they are doing. Note that this can be done with a common language, and does not require a restrictive set of rules,
* create standardized linkages between the groups so that electronic communication can be used. Note that in order for this to occur, with minimum difficulty, there must be precise rules, however these rules need not restrict the method or procedures involved.

HOW INFOODS IS PROCEEDING

To reach these goals, INFOODS is proceeding with work in three distinct areas:

I. the international linking of people and organizations to determine the status of the field and set up channels of communication;
II. the establishment of standards and guidelines to permit and promote exchange of expertise and data, designed ultimately to lead to compatibility and consistency; and
III. the development of the "machinery" to permit an actual linking of food composition data through electronic channels and media.

I. Linking Individuals and Organizations

The first task is that of linking those interested in and involved with food composition data. This began with the establishment of a secretariat office at the Massachusetts Institute of Technology, serving as the focus of INFOODS. The more important among the INFOODS secretariat activities are:

The International Directory of Food Composition Tables. We have prepared a listing of the food composition tables which are currently in use around the world. Our first edition was published over a year ago; a corrected and augmented edition is due out this month. It contains 177 tables, primarily from outside the United States. It contains the tables that we have been able to find, and we have looked long and hard. There are, however, some limitations to this directory:

* There are few US tables included (other than those of USDA)--we refer those interested to the directories of Professor Loretta Hoover of the University of Missouri-Columbia for databases and Darlene Hildebrand of the University of Washington for programs.
* We include few industrial databases (e.g., those of food companies). There are lots of these in the US, and elsewhere as well. They tend to be difficult to find out about, and more difficult to access.
* The focus is on published tables rather than electronically maintained databases. There are not too many electronic databases about at present.
* There is no easy cutoff point in the spectrum from food composition data tables to articles in the literature reporting food composition data. We have made ad hoc decisions.
* There is lots of interesting and useful information about these tables that is not included in the directory--such as what foods and nutrients each table includes. We plan to expand the directory to include much of this, as part of the production of a computer readable (and consultable) directory database.

The INFOODS Newsletter. Every three months INFOODS distributes a newsletter to everyone we know about that we think should be interested in food composition. This begins as our NCI

quarterly progress report and is massaged into something which we think will be of interest to the community. We are currently including such things as future meetings of interest and new and revised tables that we receive. Like any newsletter we are always looking for news and information to include, and ask that you send relevant items to us.

Journal of Food Composition and Analysis. By the end of the year we hope that the INFOODS Journal of Food Composition and Analysis will have begun to appear. This journal will publish articles in the areas of:

* New data and methods for food components, additives, and contaminants;
* Effects of processing, genetics, storage conditions, preparation on the composition of foods;
* Quality control procedures and standard reference materials for use in the assay of food components;
* Statistical and mathematical manipulations involved with the preparation and utilization of food composition data; and
* Processes of development and dissemination of food composition databases.

This journal is sponsored by the United Nations University, published by Academic press, and edited by Professor Kent Stewart of Virginia Polytechnic Institute and State University. We are actively soliciting papers for the first issues, and anyone with work or ideas in these areas is urged to get in touch with Professor Stewart.

Regional FOODS. In the United States, the NNDB serves as a focus for food composition database activities, and INFOODS has relied heavily on it and on the contacts made at these meetings. Elsewhere in the world there are few comparable organizations. What INFOODS is trying to do, with the major assistance of the United Nations University, is to forge links to those organizations that do exist, and organize them where they do not. Thus, there are strong groups in Scandinavia (NORFOODS) and Europe itself (EUROFOODS), and we have helped to organize NOAFOODS (in the US and Canada), ASIAFOODS and LATINFOODS (for Central and South America). Plans exist for similar groups in the Pacific (OCEANIAFOODS), in Africa (AFRICAFOODS), in North Africa and the Middle East (MENAFOODS), and in the Gulf Arab States (GULFOODS).

It is expected that these groups will work with INFOODS in:

* assessing regional problems;
* detailing regional resources, such as tables and data that exist;
* proposing regional needs and solutions;
* planning regional activities--workshops, seminars, courses;
* effecting interchange of information both within their region and with other regions;
* exploring the concept of a regional computer facility to hold regional data and produce national and special purpose tables; and
* advising INFOODS on how it can best interact and serve the region.

Standards and Guidelines

In order to develop communication and consistency we have begun to develop specific guides to assist in activities to insure that there be "good data", that these data be "carefully handled", and "well-described", and that they be "properly used". These considerations define the documents that INFOODS is preparing.

Data Quality Manual. One of INFOODS' first activity was that of commissioning Drs. David Southgate (of the ARC's Food Research Institute in the UK), and Heather Greenfield (of the University of New South Wales in Australia) to produce a revised and expanded version of David Southgate's 1974 manual on production of food composition data. This document covers in depth the procedures for collecting and analyzing food samples within the full context of putting together a food composition database. This document has been in review for over a year now

and it is hoped that it will be published late this year.

Data Compilation. Last October a group gathered in Washington to discuss the problems of "Missing Data"--how to fill in food composition data when analytic data do not exist. At the conclusion of this meeting four of us began to write a manual of how to compile food composition databases. The original scope of this document has expanded, as we try to deal with general concerns of both users and compilers. We currently have produced what we consider a solid, intermediate draft, and earnestly solicit review and suggestions.

Next, in order that food composition data be interchangeable, we are working in three areas of data description--the nutrients or components, the foods, and the data themselves.

Nutrient descriptions. Definition of food components is not as straight forward as we initially, rather naively, thought. The underlying problem is, of course, the discrepancy between biological activity and chemical specificity. We can observe what happens to individuals, and we can analyze what is in the foods that they eat--however, the correlation is often not as tight as we would like. For example, when a biologist or nutritionist talks of iron they talk in relation to anemia and that substance which, when given, ameliorates anemia. When chemists talk of iron they mean the element Fe. Nutrition is becoming more and more sensitized to how fundamentally these two concepts differ, and is now entering into a very exciting period of research--centered on "biological activity". However, this very excitement brings with it complexity and, especially now, confusion. We are working with Diane Feskanich, of the NCC as well as part of the INFOODS secretariat, to produce the first step toward dealing with these problems--a compendium of those food components that are included in major tables, with indication of which are the same and which are different. What this listing does is to distinguish those components which have the same name in various tables but in reality differ because different methods were used to measure "almost" the same thing. (For example, carbohydrate by difference and carbohydrate by sum.) This listing currently defines a standard set of almost 300 abbreviations which can be used so that table users (and computers) will be less likely to be confused. This listing has just recently been sent out for review, and we expect to issue generally a draft within a few months.

Food descriptions. One of our greatest difficulties has come from our attempts to generate a "universal food nomenclature scheme", to completely define and name foods in order that data could be easily compared from one table to another. It has become obvious that foods are so firmly embedded into individual cultures that such a nomenclature, while it may be possible in theory (and we have our doubts), is impossible in practice. The connotations of the name of a food defy literal translations and even explicit documentation. Moreover, it has become very obvious that how one looks at a food in terms of definition depends on just what one wants to do with the data. For each special use there are special needs--for example, with milk, a nutritionist might want to know what specifically is the milk that most people drink, an epidemiologist might want to know where the fodder for cows was grown, while the toxicologist might want to know how the milk was packaged and handled.

To deal with the problem of communication, INFOODS is developing a manual of "Food Description"--a project being run by Prof. Stewart Truswell (of the University of Sydney in Australia). This is an attempt to compile and define, not ways to "name" foods, but a vocabulary to describe foods--terms that can be used internationally to identify the major aspects of foods. It is expected that this document, which is currently still in draft form, will provide the context for the generators of data in their identifying of the foods that they analyze, and for the compilers in labeling their data. This is essential for intelligible communication.

Data descriptions. The third component of description of the food composition table or database is that of defining precisely what the numbers represent. Usually a single number is given which indicates some sort of average value, but sometimes indicates an upper or lower limit, and furthermore, this is often chosen only after a certain amount of "data cleaning" has taken place. Basically, what is needed is language for the information on what the numbers

themselves mean, in relation to a nutrient in a specific food. This needs to include how and when the number was derived, what it represents, and how trustworthy it is. INFOODS is just beginning to work on such a manual.

Statistical considerations. Additionally, since food composition data are numbers, it is felt that there needs to be some compendium of statistical techniques that are reasonable to use with food composition data. There is a lot of relevant statistical theory, however, this theory needs to be translated to the particular field of food composition data. Again, INFOODS is just starting to put together this document.

User's guide. Finally, it is obvious that there needs to be a User's guide. Such a manual would address the issues of choice and evaluation of food composition data and of relevant software; the interpretation of food composition data; the integration of food composition data with other sorts of data, such as requirement and consumption data; as well as the pitfalls that are so ubiquitous in the field. We have developed an outline for such a document, and if no one writes it soon, we will try to organize such an effort.

Linking Data

A long-term goal of INFOODS is that food composition data be conveniently accessible by computer. Reaching this goal requires work in two distinct areas--precise definition of what makes up food composition data and building the mechanisms for finding and moving them. Work in this area is being directed by Dr. John Klensin, a computer scientist at MIT, who is now helping direct and manage INFOODS as it focuses more and more attention on this topic.

Defining what gets moved around. Basic to computer linkage of food composition data is the precise definition of just what form data will take as they are moved from place to place. For this we have devised an "interchange format" which attaches to each piece of food composition data one or more carefully defined identifiers of that data point (which are called "tags"). For example, were we to want to send the USDA milk data cited earlier we would attach to the number "31" the following information:

* item no. 01-077, from USDA AH-8-1 (1976) milk, from a cow, *Bos taurus*, whole, 3.3% fat, fluid
* vitamin A, total; expressed in retinol equivalents derived from analyses of retinol and beta-carotene
* a mean value based on 2800 samples of butter

Much of our current work is directed to defining precisely how this information can be optimally encoded, without constraining the producer, compiler, and user of the data. This includes devising:

* a format for interchange of the data;
* nutrient tags to uniquely identify the components;
* classification tags to assist in the identification of the foods;
* data tags to describe the numbers and symbols involved; and
* precise identifiers for the origin of the data in terms of the region, country, table, and version from which they appear.

To date we have devised an interchange format, the nutrient tags, as mentioned above, are out for review, and we are in the midst of working on tags for foods and data. INFOODS Information Systems working papers describing our progress are available from the secretariat.

How the data get moved. Additionally, the Information Systems group has been working to develop the machinery for the actual moving of the data. We have started by using an existing network (BITNET) and written software to send data sets to those who make requests. This is working intermittently now, as we experiment with it, in a very prototypic mode using USDA data, and we are now exploring and defining the necessary extensions in order to get to the stage where most existing databases can be delivered to the user upon request to a single node.

The various components of such a system, as is currently envisioned, include:

* protocols for network access--machinery to utilize the capabilities of specific computer networks.
* database servers--computer programs for automatically locating and sending out data to those who request it over a network.
* conversion programs--programs to convert standard databases into and out of the interchange format.
* automated directory--programs and a database to initially inform the user of the availability of specific sets of data.
* query language--a computer language to facilitate the requesting and receiving of food composition data.

These projects and their various aspects are also detailed in the series of Information Systems working papers which INFOODS issues. These are available for comment, and such comment would be very welcome.

THE FUTURE

What is the object of all this effort? The goal is that, sometime in the future, the world will be served by a number of regional food composition groups which will coordinate and link the activities and the data within geographic regions. The data facilities associated with these regions will be linked together so that access to all the data of the world will be "easy". We are not there yet, and there are lots of difficulties in this plan, but, our ultimate goal is that:

* someone with a specific problem requiring food composition data (e.g., determination of the potential nutrient intake of a group of Southeast Asian vegetarians living in Texas) can sit down at a terminal and find exactly what relevant data exist, and then request and obtain those data in a form easily adapted to his or her needs;
* someone interested in exploring some particular aspect of food composition data (e.g., vitamin A in milk around the world) can sit down at a terminal and put together a food composition database with bits and pieces from around the world, and have confidence in the result;
* someone about to embark on an analysis program (e.g., indigenous foods of Patagonia) can easily determine what has been done already, what the best methods are, what aspects of the effort need special attention, how to best make the results available to others, and even where to turn for assistance.

For such a vision to be realized, it is obviously necessary for the tasks outlined above to be completed, and this requires the involvement of the entire food composition data community. INFOODS is now producing guidelines documents for which we solicit comments and suggestions. This is an important step but only the first.