APPLICATIONS AND AVAILABILITY OF INTERNATIONAL DATA:
MYTHS AND REALITIES

John C. Klensin and Roselyn M. Romberg

INFOODS Secretariat, Room 20A-226
Massachusetts Institute of Technology
Cambridge, Massachusetts

INFOODS Involvement and History

INFOODS was created as the consequence of a series of conclusions about the necessity and desirability of international work on, and exchange of, food composition data. Almost from the beginning, we concluded that most of this work should be done in regions of the world with similar food cultures, rather than centrally as a single international activity. As we have discussed at earlier Databank conferences,\(^1\) there has also been some doubt about the amount and quality of data available in much of the rest of the world, and it was felt that a regional approach might be more effective in developing data where none existed. During the last six years, regional groups have been organized to cover most of the world. As our earlier reports have indicated, some of them, including EUROFOODS and NORFOODS, had organizations and activities in operation before INFOODS itself was fully organized; others, such as AFROFOODS and MENAFOODS, are still in the process of holding organizational meetings. The project's initial goals and thinking were discussed in more detail in the report of the organizational meeting.\(^2\)

During its initial four years of operation, INFOODS also initiated several scientific and standardization activities to address gaps in the literature, to help focus its own efforts, and to respond to its particular mandate to provide the technical structure for efficient and unambiguous data interchange among regions. The results of that work appear in several publications, discussed below, and in plans for additional work.

Development of National and Regional Tables

There have been large changes in worldwide food composition activities over the last fifteen or twenty years. At the beginning of that period, many countries outside North America and western Europe relied on externally-compiled or regional tables, the latter often compiled by FAO in conjunction with the US government.\(^3\) Today, most countries have produced national tables or have active development efforts underway to produce them.

There is a great deal of activity: within the last six months, we have received, or received notification of, new tables from such diverse countries as Iceland\(^4\) and Chile.\(^5\) Many of these tables are exceptionally interesting relative to what we have historically expected. For example, the tables from Japan\(^6\) not only consist exclusively of original analytic values, but values are not carried forward from one edition to the next; there is no imputation of value from previous versions of the tables. The Soviet tables are also based exclusively on original analytic values, and the forthcoming edition will probably contain English translations of food and nutrient names. New tables are under development in the South Pacific area, with values for many indigenous foods, especially root crops, not reflected in earlier tables. New tables are being developed for parts of Africa, both locally and with the assistance of scientists from other countries. Even such smaller countries as Malaysia are producing their own tables,\(^7\) with apparently high analytic quality. All of these are intended as national food composition tables. If one includes the byproducts of activities such as Nutrition CRSP, which could be used as, or incorporated into, food composition tables, the list would become even longer. Many of these efforts are being strongly influenced by INFOODS efforts and recommendations for methods, terminology, and description of foods and nutrients.

Things are also changing in the computerized database area. The first generation of nutrient composition data to be released in computer-readable form, including the USDA standard reference database tapes, tended to reflect the organization and structure of the printed tables. More recent efforts have often developed database structures more suitable for active manipulation by computer, or easy transfer into other
systems. INFOODS has initiated efforts to develop regional data centers, compatible with INFOODS interchange recommendations, in East Asia, Southeast Asia, and Latin America (we have tentative approval to initiate work on the latter in early 1990). EUROFOODS has restarted its efforts to build a common table and database for the European Community and associated countries; that center should be INFOODS-compatible as well.

The irony of these developments is that, having had a leadership role in food composition work for many years, the US and North America are in serious danger of slipping behind other countries and regions. Instead of keeping up or taking the lead, the US has been indulging in complacency and intra- and inter-organizational struggles that squander the limited resources available. We have also initiated expensive and misguided efforts to create monolithic consolidated databases by forcing or bribing other countries and researchers into particular models and ill-defined purposes, rather than participating in truly cooperative international efforts.

The Regions and Their Activities

EUROFOODS held its fourth plenary meeting from 30 May through 2 June in Uppsala, Sweden. Over 90 representatives from 24 countries were present. EUROFOODS has been concentrating on improved communication within greater Europe. It was organized roughly concurrently with INFOODS, and its first major meeting was held earlier. The meeting included reports of several new and important scientific developments in food composition work carried out under EUROFOODS auspices, in addition to reports of continuing work of general interest. The proceedings of the EUROFOODS conference will be published as a special issue of the INFOODS Newsletter, probably early in the last quarter of this year.

NORFOODS works closely with EUROFOODS, and most of its members participate in both. At the same time, NORFOODS -- an official activity of the Nordic Council of Ministers -- meets regularly and carries out activities of mutual interest to improve food studies and food composition data and their interchange within the Nordic countries. The governmental support and long traditions of cooperation make NORFOODS probably the best organized and most active of the regional groups.

OCEANIAFOODS held its first full conference in Canberra, Australia, in May, 1987. Another meeting is planned for the near future. A major effort is underway to prepare new tables, based primarily on original data, for the South Pacific Commission. New tables, built on INFOODS recommendations, are being completed in New Zealand, and a new Australian database has been released.

LATINFOODS, has held several informal meetings and a two formal ones, the most recent in Chile last November. LATINFOODS is reaching agreement on standards for sample identification and table construction, and is working toward the development of a regional data center.

AFROFOODS, including Africa south of the Sahara, held an organizational meeting in September, 1988, at the Third Africa Food and Nutrition Congress in Harare, Zimbabwe. Plans are being made, and resources sought, for a conference later this year or early in 1990.

ASIAFOODS, covering East Asia, South Asia, and Southeast Asia, held an organizational meeting in 1984. Additional meetings of subgroups have been held in conjunction with nutritional congresses in the region. Discussions are underway that may lead to the development of regional data centers in Asia during the next few years.

MENAFOODS began to organize this year, with participation from North African and Middle Eastern countries. Commitments have been received from several countries, and an organizing committee selected.

NOAFOODS, covering the English-and French-speaking portions of North America and the Caribbean, is largely dormant, and is likely to remain so until resources and a location are identified for a regional data
center. Annual "National Nutrient Databank" meetings in the USA continue to bring experts and users of food composition data together for discussions of topics of interest and serve some of the functions of a regional organization.9

Publications and New Results

(1) Users and Needs: INFOODS sponsored a conference in March, 1985, to provide a better understanding of the ways in which food composition data were used, and the problems associated in storing, organizing, and processing them. This conference in many respects followed up the organizational meeting, focusing on these specific topics. The results of the conference were summarized in a book10 published two years ago.

(2) Index and directory efforts: For the scientist in search of data values from outside his or her own country or region, one of the most important tools is a good directory of what tables are available. FAO used to produce such directories,11 but stopped doing so. More recently, in-depth directories have been produced in a few regions, notably here in Europe12 and in North America.13 INFOODS has supplemented these, and provided listings for regions without their own more detailed listings, with an international directory. That directory has been published in two editions, the more recent one last September,14 and additional editions will be produced when the volume of new tables and corrections requires them. Resources are being sought for a more ambitious, and more computer-intensive, project that would provide an international index of food composition data and databases, organized by foods and nutrients, not only by geography.

(3) Food component identification: Somewhat unexpectedly, differences among countries and their tables about which nutrients and other food components are included in tables and how these are analyzed raise major challenges to comparison of results that are based on different tables. For example, some countries report only carotene for "vitamin A". Others report carotene and retinol, while still others report retinol equivalents, and so forth.15 Misunderstandings occur when these values are compared as if they were all the same. INFOODS first surveyed this problem to determine its extent and to understand how precise various food tables were about identifying the methods used and chemicals being reported. The problem is quite extensive, and, not surprisingly, tables differ as significantly in their degree of precision and detail about nutrient and method identification as they do about other element of data quality. A system of identifying specific foods components was then developed using a great deal of international participation and comment.16 While that system was designed primarily for data interchange purposes, it may also be useful for organizing tables and reporting methods of analysis and calculation in printed tables. Several countries have already begun to use it in these ways.

(4) Analysis methods and table compilation: Much of the complexity and detail of the food component identification system stems from historical differences in procedures and approaches. If we could start over today, rather than using older tables and data, and could agree on a single set of standards and conventions, many of the problems would be eliminated. Unfortunately, starting over is unrealistic and will be for some time: there is a great deal of very useful older data. However, for new tables and databases, and for extensive revisions of existing tables, determining how to do things well and in a generally-agreed-upon fashion is more desirable than simply describing what was done in the past without making normative judgments about those practices. INFOODS initiated two efforts to provide guidance for the developers of new tables and databases. The first of these, concerned primarily with methods of food selection, sampling and analysis, has culminated in a manuscript by Heather Greenfield and David Southgate. The second concentrates more on the organization of food composition tables and databases for specific tasks and on ways of estimating values that have not been directly measured. Both are expected to be published before the end of this year.

(5) Data Interchange: The data interchange system itself is designed to provide a standard format for transferring food composition data among regions. It may be useful as well for data transfer and organization within regions and even within countries. The system is very precise and permits storing and transferring any
data, or information about the data, that are available without need for the separate introductions, usage notes, and ancillary files that have characterized formats and database organizations in the past.

Preparing a file for interchange involves converting it to a special structured format in which all data, and all descriptions of data, are specifically and precisely identified with special names or "tags". Unlike historically more common data formats, this permits whatever information is associated with the file to be represented and exchanged, without requiring extra space or data fields for information -- whether nutrients or description -- that is not available for the file. The specific details of the interchange system have been explained at conferences and are covered in a series of working papers that have been widely circulated. Those working papers and some new results have been consolidated into a lengthy technical manuscript that includes rules for syntax as well as all of the tags that are not associated with food component identification. We expect to publish this manuscript when all of the ideas can be tested together and in context, i.e., in conjunction with the implementation of the first regional data center.

(6) A forum for scientific papers: Finally, the INFOODS-sponsored Journal of Food Composition and Analysis published its first issue in late 1987, and has been publishing issues at regular intervals. It is now operating to a great extent independently of the Secretariat. Subscriptions are available through Academic Press.

Goals and Progress

From one perspective, INFOODS has had three goals:

(i) Improving communication among scientists interested in food composition data and organizing effective regional groups. Partially because the time was right for regional development, and in a few cases without significant input from INFOODS staff, this process has been very successful. A periodic INFOODS newsletter and the Journal of Food Composition and Analysis also contribute to improved communications.

(ii) Identifying the scientific barriers to effective inter-country and inter-regional use of food composition data, and eliminating or lowering those barriers. Significant progress has been made, especially in the areas of guidelines, data identification, and facilities for meaningful data interchange. In other areas, much work still remains to be done, either to solve problems or at least to increase the general understanding of them. The area of food coding and nomenclature is especially important: many systems have been proposed as meeting all needs and being applicable for broad international use, and the usual assumption is that precise food identification is a problem that must be solved. These assumptions have led both to good systems and serious fallacies, discussed in more detail below.

(iii) Providing for actual international data availability and exchange. Many of the prerequisite problems in this area are now solved, as mentioned above: An up-to-date directory of tables has been completed, and several regions, including EUROFOODS, have developed comprehensive surveys of data available in their areas. An interchange system has been designed and tested with data from several tables from various parts of the world. The next step, and the one on which INFOODS is concentrating its efforts, involves the development and construction of operational regional data centers that can retain and exchange data and, where appropriate, act as a focus for specifically data-related activities within their regions.

Status of the Secretariat

With the completion of the initial specific INFOODS core tasks and expiration of the initial core funding, the INFOODS Secretariat has entered a period of consolidation and reorganization, focusing its efforts on obtaining the resources to construct regional centers in a few parts of the world. The proposals and discussions so far assume joint efforts, with the Secretariat providing assistance and technical expertise, but with most of the actual development work being done within the target regions, by regional personnel.
Fund-raising efforts so far have concentrated on regions in developing areas, partially on the assumption that the more developed areas will eventually take care of themselves; however, we are looking for independent or partnership arrangements in developed areas as well.

Partially as a result of this shift in emphasis, visible Secretariat activities have been considerably reduced in scale. The Secretariat does, however, continue to represent INFOODS at regional meetings, including all of the regional meetings mentioned above, and, when resources permit, at important scientific activities related to food composition. For example, the Secretariat has organized an INFOODS session for the nutrition meetings in Seoul this August; I hope to see some of you there. The Newsletter will be published less frequently and less regularly, at least until we have more to report, and incoming letters and requests are taking longer to be answered. We apologize for any inconvenience, but these delays should not be taken as an indication that we have disappeared; we have not.

What can be Done with International Data?

All of the work and organizational effort we have described provide mechanisms and incentives for improvements in food composition data in individual countries and regions and provides the basic structure for efficient data interchange among them. These arrangements provide an affirmative answer to the obvious question of "can the data be moved?". Once the regional data centers are in place, we should have a positive answer to "can it be done smoothly and efficiently?" as well. Although, unless there is some activity in North America, the answer could be "no" here and "yes" in the rest of the world. However, in some respects, questions have been answered out of order; the more serious questions of "is it of any use?", "can the data be interpreted?", "is it possible to match foods from different tables?" have so far remained unanswered. There are tentative answers for some of these questions. For others, there is even less certainty. The controlled comparative studies that are needed to produce definitive answers have not been done, and it appears that there is no interest in supporting the required research.

These fundamental questions have been blocked by the focus on food names.17 The belief that, if only all food composition tables used the same naming system, then all other problems would disappear would, if true, eliminate the need to answer them. Not only are there obvious reasons why this is not the case — naming foods does not relate, for example, to comparing nutrients measured in widely different ways — but it has become unclear that food naming, per se, is even a worthwhile question for investigation.

Food Terminology and Classification: The key question or an interesting distraction?

This food-naming problem and its implications are particularly acute where international or cross-cultural data are of interest. It has been the most controversial of the questions INFOODS has had to deal with, in part because many people seem to have strongly-held opinions about it. The INFOODS conclusions are perhaps somewhat radical, and treated in more length below.

The question of food names, coding, and description, is key to many ideas for comparing values between different tables and databases, as well as to food coding for dietary studies. The idea of developing a single international nomenclature system, on which everyone can standardize, is quite seductive. Unfortunately, it does not appear to be practical, and regional systems that focus on particular goals and objectives seem to be more reasonable. Systems such as the EUROFOODS-developed EUROCODE 210 have precisely the correct set of properties:

The code's coverage reflects the foods of a relatively small number of countries, with fairly similar foods.

The goals for use of the system are well-defined and limited.
As one moves away from this set of properties, systems inevitably become either more complex and less problem-specific or very general and imprecise. They may even become unworkable. This is not a defect in any particular system that is subject to improvement by adding or changing a few terms; it is inherent in the nature of the underlying foods and varying societal assumptions about them. This is perhaps best illustrated by a series of examples.

(1) For dietary assessment purposes, a food coding system should reflect only those distinctions among products that the consumer can make in the marketplace. Additional distinctions and the attempt to elicit them in a survey are, at worst, sources of scientific error and, at best, merely a waste of time and classification effort. By contrast, a regulatory database may require considerable additional classification information and coding; organizing principles based only on consumer distinctions might not be useful. Adequate computer programming would permit use of the less-precise parts of a very detailed system for dietary assessment without other effects, but only if the detailed system was designed using appropriate criteria for a dietary assessment system. No one has yet demonstrated that this can be done optimally for a general-purpose food composition database.

(2) A multilingual thesaurus is a useful supplement to a food coding system and such thesauri have been constructed, primarily for agricultural products generally, but also for human ones. However, once a system is adapted to regions with widely differing food cultures, a simple thesaurus, which translates words of one language into words of another, is no longer sufficient: it is necessary to describe the foods in great detail, comparing them to local products when they exist. In the terminology of thesaurus systems, this implies carefully-translated scope notes, not merely translation of terms.

(3) Some of the food similarities implied by the ability to create multilingual (and multicultural) scope notes simply do not exist, i.e., there are no corresponding foods in different countries. Similarly, marketplace distinctions in one country may be different than those in others: more precise in one place, less in others. One can easily invent a system that permits local variations to compensate for this, but then one has a common method for building coding systems with similar features, not a single international system.

An example from another area that many of you have encountered may help to illustrate the underlying problem. Since computer systems began to become important for processing things besides numbers, there has been a concern about eliminating the tyranny of character sets, based on US English, in which most European -- and other -- languages could not be correctly expressed. After a slow start, there has been considerable development in registration of national graphic character sets and devices to handle them. If the registration sequences are used, one may still not be able to use a French character set on a US-developed device, but at least no assumption will be made that it is "pure" ASCII. At the same time, the hardware and software technology now permits more convenient use of multiple or extended character sets: many of you have noticed that INFOODS has gradually succeeded in spelling your names and organization names correctly, something we could do only by marking in letters that did not appear in US English with a pen just a few years ago.

(4) In some cases, where the differences between marketplace distinctions in one country and those in another are only a higher degree of distinction in one than the other -- for example, it is well-known that Denmark and France identify far more distinct cheeses than the USA -- it appears possible to build an international system by adding hierarchical levels for those countries which will use them, with others ignoring those levels. This is useful, practical, and important, as long as its implications are understood: comparisons between the foods of those countries can be carried out only at the minimum of the level of distinction made in either. There is no way that a coding system will invent information that does not exist, and it is improbable that the use of such a system will, of itself, change what manufacturers sell in the marketplace.

(5) More serious problems arise when one group of countries have different, non-overlapping systems for naming or selling foods. The most commonly cited example in this area is that of cuts of meat, but there are others.
(6) And, of course, while no coding system can eliminate simple errors, such as the identification of a highly processed cheese product as "cheddar", more complex and precise systems may be more prone to errors than simpler systems in single countries, where terminology is more obvious. Description of foods in prose can act as a useful check on complex classifications and can be used to resolve or approximate differences among them. If the "cheddar" above were also described as a "soft, processed cheese product with several ingredients", then the potential for serious scientific error when comparing values in multiple tables would be reduced. On the other hand, if that description were there, it is not clear that assigning a specific name to the product adds much information, at least outside the country that produced the food table in question.

In summary, these difficulties, especially the potential for errors of classification and the lack of precisely analogous foods that makes multilingual thesauri very difficult, are strong arguments that the "general food code" problem should not be the primary focus of international efforts. Instead, efforts should concentrate on the development, where needed, of highly-specific problem-oriented codes and the development of good food description procedures and checklists for more general comparison and exchange of values. If that approach is taken, the unsolvable problem of a universal food nomenclature will become moot.

REFERENCES


5. Tabla de Composición Química de Alimentos Chilenos and Materiales para consumo habitual y potencial en Chile, Santiago: Departamento de Ciencias de los Alimentos y Tecnología Química, 1985.


