JOURNAL OF FOOD COMPOSITION AND ANALYSIS (2000) **13**, 709–744
doi.10.1006/jfca.2000.0891
Available online at http://www.idealibrary.com on IDEAL®

REPORT

EUROFOODS Recommendations for Food Composition Database Management and Data Interchange

Florian Schlotke*[,1], Wulf Becker†, Jayne Ireland‡, Anders Møller§, Marja-Leena Ovaskainen¶, Judit Monspart‖, and Ian Unwin**

**Institute of Scientific Computing, ETH Zurich, Switzerland; †National Food Administration, Uppsala, Sweden; ‡AFSSA-CIQUAL, Paris, France; §Danish Veterinary and Food Administration, Søborg, Denmark; ¶National Public Health Institute, Helsinki, Finland; ‖Department of Canning Techniques, University of Budapest, Hungary; and **Food Information Consultancy, Cambridge, UK*

Received July 23, 1999, and in revised form February 14, 2000

To promote and encourage interchange of food composition data, the EUROFOODS working group on food composition data management and interchange proposes a set of recommendations for data management and interchange using electronic media. The recommendations are firmly founded on previous work done internationally by INFOODS and by national agencies and institutes as well as international standards. The recommendations include guidelines for the description of foods, components, compositional values and data sources. A sufficiently generic conceptual schema for food composition is defined to handle food composition data at various levels of aggregation and with various levels of additional descriptive information. The recommendations also include technical issues such as file formats and media for data interchange. Software tools are presented to assist with implementation of the recommendations.

Key Words: food composition data management; data interchange; data quality; metadata; food description; EUROFOODS

1. INTRODUCTION

1.1. *Motivation*

The results of a questionnaire that was sent to 25 European countries to compile an inventory of European food composition databases (Møller and Schlotke, 1996) led to two main conclusions: first, systematic electronic interchange of food composition data is important, because data compilers depend on data from different sources or like to share experience and resources. Second, data interchange among data compilers is most wanted at earlier stages of data production, i.e., levels one and two according to the four-level production framework proposed by Greenfield and Southgate (1992).

Despite this need, data are currently not interchanged systematically on the international level due to the following problems (Schlotke, 1996): Interchange is mostly done *ad hoc* or an a bilateral basis only (*organizational aspect*). Interchange is not or seldom formalized: different software and file formats are used; data are not sufficiently described and are therefore often hard to interpret correctly and

[1] To whom correspondence and reprint requests should be adderessed.

0889–1575/00/040709 + 36 $35.00/0

unambiguously (*logical aspect*). Interchange is often doen on paper only or using computer media that cannot be read by the person receiving the data (*physical aspect*). Interchange is hindered by copyright constraints (Ricketson, 1995) (*legal aspect*).

Consequently, it was decided to develop a set of recommendations and tools to promote data interchange in Europe. This paper focuses on recommendations that address the logical and physical aspects of data interchange. If an interchange system supports these aspects, organizational networks and copyright policies can evolve.

1.2. *Objectives and Scope*

The above observations lead to four general objectives to be met by the proposed management and interchange system.

Generality: to promote and encourage active electronic interchange of food composition data at all stages of the compilation process. This includes data interchange among data producers (i.e. laboratories), compilers and users in Europe and beyond. Therefore, the system should be generic and able to cope with the data involved at any of these stages (e.g., raw and aggregated data). Note that in this context food composition data include compositional, as well as qualitative, information about foods.

Completeness: to encourage the collection, electronic storage, and interchange of sufficient metadata to describe and identify food composition data. Plain figures are meaningless as such. Food composition data must be sufficiently documented for proper interpretation and usage. The additional data needed to describe the actual data, its nature and production state, are referred to as metadata. In our context, metadata include source-, food-, component-, compositional value and method-description and can be found in carefully prepared scientific papers and laboratory reports.

Flexibility: to allow for easy addition of new types of metadata. The types of metadata can never be fixed at any point in time by some authority. Therefore, the system should allow storing and interchange of all metadata available for a given data source, even if some types of metadata are not standardized today but might be relevant for future applications. Such additions should be possible without causing major reprogramming of the system.

Implementation friendliness: to allow for implementation of the system with reasonable effort using established technology (e.g. relational database management systems or spreadsheet applications). Software tools must be provided to facilitate and ease data interchange and management. The tools should allow the transfer of data between the interchange system and locally used food composition database management systems or any other widely used software package (e.g. spreadsheets, statistical software, etc.).

This paper presents a set of recommendations for data management as well as data interchange. The focus, however, will be on data interchange issues, as it is not intended to interfere with existing data handling procedures at the various data centres. The recommendations do not serve as a fixed set of rules. Therefore, when applying the recommendations, the user is free to

- Extend the recommendations with new rules addressing issues not yet covered by the recommendations. Such extensions should not affect the interpretability of those parts that follow the recommendations. Any extension must be documented and the documentation must be accessible for the data receiver.
- Implement only part of the recommendations; in which case the invention of new solutions for issues already covered by the recommendations must be avoided. In any case, the recommendations suggest a minimum set of information that must be provided in data interchange.

Finally, it should be noted that implementing the recommendations does not in itself imply any degree of quality assurance of the data. Quality assurance is part of the data description, the metadata. On the other hand, the recommendations allow the receiver or user to interpret the data in a regular and standardized manner and to judge the data quality based on their intended use.

1.3. *State of the Art in Food Composition Data Interchange*

Many contributions served as a basis for the recommendations proposed in this paper. One part covers technical specifications for data management and interchange. Other work has been done in various directions to harmonize terminology and procedures in such fields as food-, component, method-, value and source-description.

1.3.1. *Technical issues.* Several national food composition programs defined precise technical specifications for the publication of their food composition tables (BGVV, 1996; Burlingame *et al.*, 1996; Corkill, 1995; Favier *et al.*, 1995; Møller and Saxholt, 1996; USDA, 1998). Primarily designed for end-users and also published in printed form, these tables offer only limited additional description of data and the level of detail is generally not sufficiently specific to be used as input by compilers in other countries. The format and content of these tables is mostly incompatible. Nevertheless, they served as a rich source of ideas for the more general recommendations presented in this paper. Two initiatives have been reported on the international level to harmonize data formats. First, the NORFOODS Computer Working Group discussed and practised data interchange among the Nordic Countries from 1985 onwards (Møller, 1992). Although a lot has changed in computer technology (e.g. networks) since then, this work has shown that data interchange is possible with only a few straightforward rules that are easy to implement with respect to the finance and skills involved. But the group also mentioned that data interchange would be easier if data files were more alike in terms of format, terminology and further documentation of the data. Second, the INFOODS organization always considered international food data interchange as one of their primary goals. Between 1986 and 1992, three sets of recommendations have been published: a system for food component description, the so-called *tag-names* (Klensin *et al.*, 1989), a framework for food description (Truswell *et al.*, 1991), and a data interchange format (Klensin, 1992). Up to now, only the tag-name system has been implemented and is used by a number of agencies world-wide. The INFOODS data interchange system has not yet had much success. The main reasons are the lack of software tools that support this format and a conceptual problem of the format that makes it hard to write these software tools in practice. These problems are discussed in detail by Unwin and Møller (1996). The ongoing project "Food Table Viewer" software by Unwin (1999) has provided further experience with data management and a mechanism for practical and data interchange, especially at the level of data from published food composition tables. This project also addresses the question of metadata and its harmonization. The project continues with the opportunity for wider collaboration and contribution.

1.3.2. *Food description.* Food description includes food names, food classification, sampling procedures and information on food properties such as food source, agricultural production and storage conditions, preservation and cooking methods, food additives, etc. More than 50 properties that influence the nutritional value of a food have been identified (Truswell *et al.*, 1991; Pennington *et al.*, 1995). Pictures are also a possible way to describe foods (Burlingame *et al.*, 1995). An overview of recent work in food description can be found in Pennington (1996) and Ireland and Møller (2000).

Besides plain textual description and pictures, there are basically three techniques used for food description:

1. *Monohierarchical classification systems* like Eurocode 2 (Kohlmeier, 1995), the CIAA Food Categorization (Codex, 1995) or the numerous proprietary food grouping systems used in each country. Although single classification systems are powerful tools within specific application domains, they cannot cover all relevant descriptive information needed in food composition data assessment. Such classifications organize foods according to only one property (e.g. biological origin, nutrient content or legal aspects). In most cases, more than one property needs to be described in order to get a sufficiently detailed picture of a given food (Truswell *et al*., 1991). Another problem with monohierarchical classification systems is that for each food (or type of food) a distinct slot within the hierarchy needs to be defined and fixed forever at design time. This can lead to inflexible and huge classifications. A practical problem arises when designing classifications for international use: in different cultures, people see relationships between foods in different ways. A consensus on a fixed classification is often hard to achieve at the international level.
2. *Faceted description systems using standardized vocabularies* (*thesauri*): To overcome the inflexibility of monohierarchical classification systems, multifaceted food description systems have been developed. A given food is described with respect to several facets (i.e. viewpoint, properties or attributes). An example is the LanguaL system, originally proposed by the United States Food and Drug Administration (FDA), with its 14 facets (Hendricks, 1992; Møller and Ireland, 2000). For each facet, a standardized vocabulary (i.e. a set of possible terms or descriptors that may be applied) is defined in a thesaurus. A unique alphanumeric code is assigned to each descriptor. These codes can be used for international data interchange. LanguaL is currently maintained and extended under the COST Action 99—EUROFOODS initiative.
3. *Faceted description systems using free text*: This approach was proposed by the INFOODS working group on food description, terminology and nomenclature (Truswell *et al*., 1991). This system differs from LanguaL in the sense that far more facets are proposed (about 50) and not all facets are supported with standardized vocabulary. Generally, free text can be applied to describe a given food with respect to a given facet.

It is not worthwhile arguing which of these systems is the "best". Each system has its specific purpose, and it has advantages or disadvantages under different conditions. As a result, features from the different systems can be implemented together. For example, Eurocode 2, the German BLS-code and the Slovakian faceted food code, mix the concept of a hierarchical classification with the faceted approach. The strength of classifications and LanguaL is their strict definition of vocabulary and usage of codes, which makes these systems language independent (but not necessarily culturally independent) and suitable for systematic computer processing. The INFOODS system, on the other hand, is much more flexible but with the price of being less formalised which can lead to misunderstandings in data interchange and imposes difficulties on computer-based data handling. As a conclusion, all three techniques, and others like the description of foods using pictures, should be used to complement each other. Such a combination of approaches was proposed by an FDA initiative called International Interface Standard (Pennington *et al*., 1995).

1.3.3. *Component description*. Component description includes information on the type of component, the methods used to obtain compositional values, the units and the modes of expressions used to express compositional values (Klensin, 1992; Unwin and Becker, 1996). INFOODS developed a list of standard abbreviations for components to be used in data interchange. This list of so-called tag-names evolved out of a survey of components found in major food composition tables world-wide. Information on component description (component name, unit mode of expression and in some cases method of analysis or derivation) is part of the definition of each tag-name. Components found in different food composition tables but using the same tag-name can therefore be considered to be compatible. The INFOODS tag-names are used at an increasing number of agencies throughout the world and help users to compare published food composition tables. This approach, however, has several disadvantages when used at earlier stages of data compilation: a food database compiler often needs more information than is covered by the INFOODS tag-names (e.g. accuracy of the method used). The tag-name is inflexible, especially when dealing with components whose definitions depend on various analytical methods (e.g. folates). Each new combination of the various aspects needs a new tag-name to be registered. It is easier to manage several more stable collections of standardized terms for each aspect of component description, than one list of tag-names representing many combinations of the basic terms. A more practical problem is that not all tag-names representing many combinations of the basic terms. A more practical problem is that not all tag-names are described with a method (and mode of expression). It is argued that these components are *rational* in the sense that the compositional value is independent of the (presumed) analytical method used. In this respect, the tag-name system implies a preliminary judgement of whether two components are compatible. This might be useful for the lay user but not for the expert compiler who is interested in more "raw" data. In contrast to the INFOODS approach, the component aspect identifier system (CAId) suggests separating the various aspects of component description.

1.3.4. *Method description.* Method description includes analytical as well as computational methods to generate food composition values. A proposal for harmonization of such descriptions is given in the CAId system (Unwin and Becker, 1996). Another source of information on analytical method description is the Codex Committee on methods of analysis and sampling (Codex, 1997b).

Value description: Value description documents the expected variability of a compositional value and includes data on the statistical distribution of analytical measurements and indication of values that are missing, below detection limit, trace, etc. Value description is discussed in the INFOODS data interchange handbook (Klensin, 1992). In practice, however, this information is seldom managed systematically, if at all. In particular, the statistical aspect of nutrient composition has not had much attention in data interchange in the past (Klensin, 1995). The description and meaning of the terms *trace*, *zero* and *missing value* is not used uniformly in the literature (Stewart, 1988; Klensin, 1992). A proposal for standard codes to indicate the type of missing value can be found in the work of NORFOODS (Møller, 1992). There is also some confusion regarding what kinds of information should be modelled as value description and what as component description, since they sometimes overlap (Unwin and Becker, 1996). More conceptual work and clarification is needed in this field.

1.3.5. *Source description.* Source description includes all information needed to track the sources from which food composition data were obtained (laboratory, literature, etc.). Source description of complete data files has been formalized within the INFOODS data interchange system (Klensin, 1992). It includes information about the institution and/or person responsible for the content of an interchange file (i.e. the *source*) as well as information about the person acting as the *sender* of the file. The

INFOODS system also introduced the concept that each interchange file must have just one source authority attached to it. This does not imply that all the data must come originally out of the same laboratory, or even the same country. Instead it recognizes that the activity of putting together a database involves editorial and scientific judgement rather than mechanical concatenation of values. Source information for individual values is covered in the CAId system (Unwin and Becker, 1996): a source type indicates the general category of a source such as food table, journal article, laboratory report, etc. Depending on the type of source, different types of reference information are given (e.g. bibliographic references).

1.4. *Conclusion*

Today, no standardized and comprehensive international system for food composition data interchange is in use. Most of the proposed solutions focus on the distribution of published food table data to end-users. These systems are too restrictive to be used at an earlier stage of the compilation process, where more detailed information is needed.

The remainder of this paper is structured as follows. Section 2 gives the rationale behind the recommendations and defines the terminology used throughout this paper. Section 3 covers the four major recommendations: conceptual data schema, standardized vocabulary, file formats and media for data interchange. Software tools for implementation of the recommendations are presented in Section 4. Section 5 discusses limitations of the current design and suggests further research and development.

2. A REFERENCE MODEL FOR FOOD COMPOSITION DATA

This section gives the conceptual background to the actual technical recommendations given in section 3. We propose a reference model for food composition data, which serves as a framework for both data management and data interchange. The model consists of two parts: an organizational framework and a reference data schema. The data schema is static to some extent, but allows flexible extensions for individual use. It is a conceptual schema and does not imply any specific file format or database implementation. It serves, however, as a common ground for discussion of the development of specific implementations.

2.1. *Organizational Framework: Data Management and Interchange*

Data management and interchange are closely related: both tasks handle the same information and interact with each other. The operation and technical constraints, however, are different. Figure 1 outlines different parties involved in the production chain of food composition data and typical interactions between these parties. Based on Greenfield and Southgate (1992), food composition data is managed at four levels during the compilation process:

Level 1. *Data sources*: published and unpublished research papers and laboratory reports containing analytical data. Data might be systematically managed within a laboratory information management system (LIMS).

Level 2. *Archival data*: written, printed, microfiche or computer files that hold all original data expressed as they were originally published or recorded, scrutinized only for consistency in data format. This editing process might include translation of

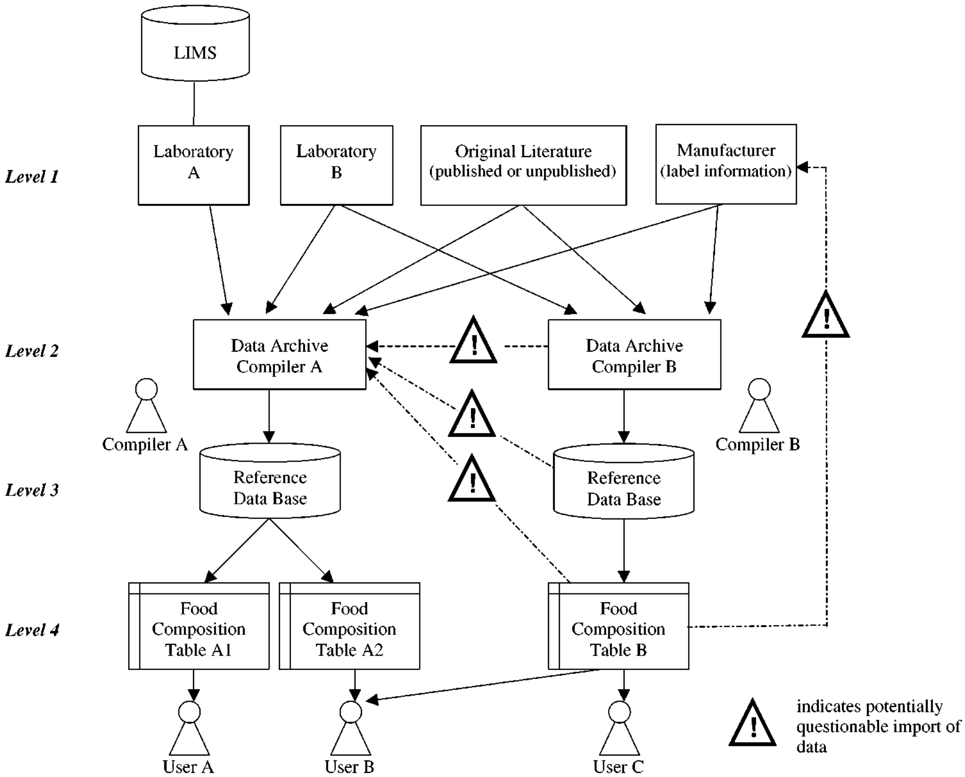

FIGURE 1. Data interchange and management at various levels.

information into standard coding or naming schemes. Such files should contain enough background information (metadata) so that it is unnecessary to refer back to the original sources. Archival data are kept by the compiler for backup purposes.

Level 3. *References database*: the complete pool of rigorously scrutinized data in which all values have been converted into standard units and components are expressed uniformly, but in which data for individual analyses are held separately. This database includes all foods and components for which data are available, and is linked to auxiliary records which indicate methods, sampling procedures, bibliographic references, laboratory of origin, date of insertion and other information relevant to the compilation process. This database can be part of a (relational) computerized food composition database management system (FDBMS). It is from this database and its programmes that the user databases and tables can be prepared.

Level 4. *User databases and tables*: the public resources which hold evaluated food composition data that, in some cases, have been weighted or averaged to ensure that the values are representative of the foods in terms of the use intended. User databases are subsets or derivations of the reference database, specially designed to meet the needs in terms of form and content of different user groups. These databases include as many foods and components as possible, with preference being given to completed data sets. Data may be completed by calculation or estimation.

Note: There is a risk that compiler A uses data from compiler B that originates from sources already used by compiler A. Since it is often hard or even impossible to trace

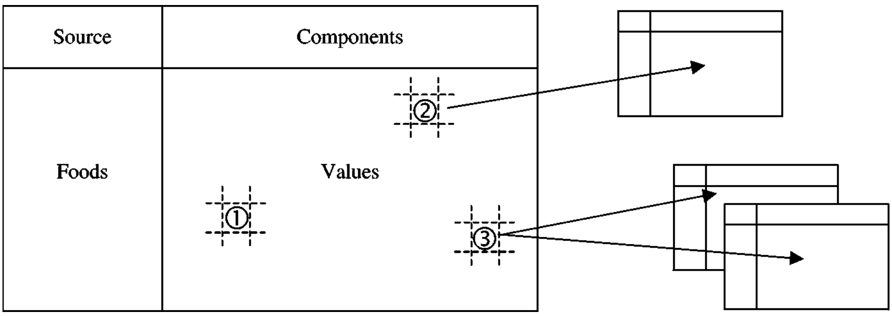

FIGURE 2. The table metaphor.

the history of data at levels 2–4, data from these levels must be carefully evaluated when used as input by an archival database compiler. A similar risk lies in manufacturer—or labelling—data of food products since these data may also have been derived from published food composition tables. Therefore, the data interchange system must enable inclusion of the contributing values and their description within the metadata of a derived value.

Based on this framework, the following definitions can be given:

Data management: any systematic form of organizing food composition data at a distinct plate, e.g., laboratory, food table compiler, food table user.

Data interchange: transfer of data between a sending party and one or many receiving parties without loss of information, i.e. the receiver should be able to interpret the data in the same way as was intended by the sender.

Interchange package: Data are always interchanged within a self-contained *interchange package* holding all the information needed to assess the scientific quality of the data. The term *interchange package* is used in a general sense without implying specific implementation techniques such as single markup files, databases or a collection of several files of various types. Specific recommendations for implementation are given in Section 3.3.

2.2. *General Data Schema*

People are used to publishing and reading food composition data in tabular form. Data are typically presented with foods in the rows and components in the columns (see Figure 2). The upper left quadrant of the table may be used to hold the information that describes the table as a whole, e.g., information about the body that is responsible for the content of the table. The foods, components and values quadrants also hold additional descriptive information on these items. Figure 3 depicts a translation from the table metaphor into the entity relationship model (ERM). A data *source* (i.e. a food composition table/study) consists of several *foods* and several *components*. Each food-component pair may yield a compositional *value*. Each *value* is linked to its *method* description. There are three basic types of values (see Figure 2):

1. a value may be an original analytical, calculated, or estimated value of this particular data source,
2. a value might be drawn from a third-party source,
3. a value might be an aggregate of several other values, which in turn may point to third-party data sources.

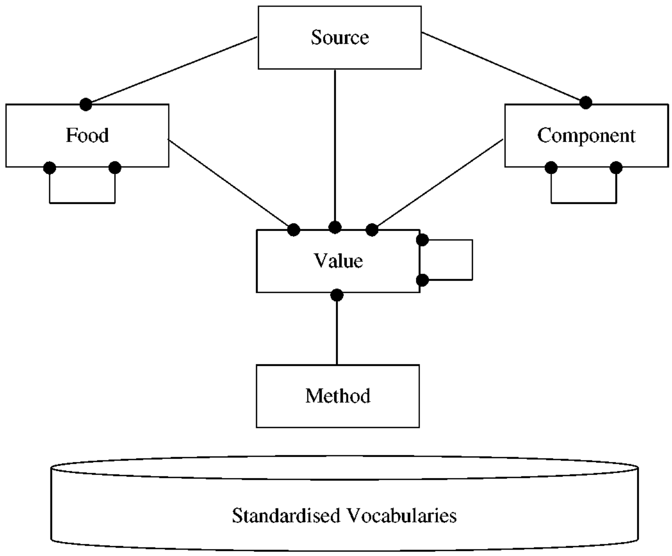

FIGURE 3. Basic entity relationship data schema.

The second case can be modelled by linking the value to a *source* entity. The third case is handled by linking the value to all its contributing values, which in turn are *value* entities.

The schema is static in the sense that the main entity sets do not need to be changed to capture food composition data at the various levels of composition. It is flexible because it provides an open framework for further metadata to describe foods, components, methods, values and data sources. A list of mandatory and optional attributes to be used in interchange packages is given in Section 3.1. Since many of the metadata attributes depend on standardized terminology, a repository for standardized terminology is part of the general data schema. Both the list of attributes and the standard vocabulary are open to future extensions.

The main entity sets are defined as follows:

Source, primary source and secondary source: A data *source* is a set of compositional values reported by a single person, group of authors or organization. This authorship takes the responsibility for the content of a *source*. Besides the authorship a single person, group or organization acts as the sender of a *source*. The sender is responsible for the formal correctness and electronic transcription of a data source. Examples of *sources* are laboratory reports, scientific papers on specific studies, compiled analytical data of specific food groups and/or components, comprehensive food composition tables, manufacturer and labelling data, etc. A *source* may be available in various forms: published or unpublished reports, journal papers, articles in books, labels, etc. A *source* must be described with sufficient bibliographic reference information in order to be uniquely identified. The *primary source* within an *interchange package* is the *source* to be interchanged with that package. *Secondary*, *tertiary*, *etc*., *sources* are *sources* on which the *primary source* is directly or indirectly based. In case of an original work, no secondary sources are needed.

Food: Within the proposed food composition data interchange system, we consider every food reported in a *source* as a single entity *food*, since no two foods or food samples reported are exactly the same. This also applies for generic foods (i.e. a representation of a class of foods that can be considered the same under a given context, e.g. "apple" in a national food composition table), since we cannot assume that any two compilers of such generic foods intend to express the same thing. Examples of *foods* are specific samples analysed in a laboratory, food products from a specific producer, generic foods and products, mixed foods and dishes. Within a data *source*, each *food* must be assigned a unique ID (e.g. a number). Even though two reported foods (e.g. two samples) might be described using identical descriptors, they are treated as two individual entities. Whether two reported *foods* are comparable and might be aggregated at a later time is a decision of the data user and depends on the application and its constraints regarding data quality. The more metadata that are available to describe the food, the more precise the decision of the user (e.g. a national data compiler).

Component: We apply the same philosophy to *components* as we did for *foods*. Each *component* reported in a data *source* is unique and must be evaluated according to the available metadata. In that sense, every distinct set of values for the attributes component name, unit and mode of expression must be considered a *component*. *Components* include all properties of food that are the subject of scientific measurements to determine the amount of property per some amount of food (e.g. per 100 g food). In particular, *components* are not restricted to nutritionally significant properties of foods. Examples of *components* are nutrients such as fats, proteins, carbohydrates, vitamins, minerals, and also contaminants or physical properties such as density, per cent edible portion or pH. Food-specific factors to be used in calculations may also be modelled as components (e.g. nitrogen conversion factors for protein calculation). All other properties of food that are not included in this definition are treated as part of the food description.

Value: A numerical result and its statistical properties determined by an analytical process, computation or estimation of the amount of a *component* within a *food*.

Method: Chemical, physical or numerical methods to determine *values* of *components* within *foods* as reported in *sources*.

Standardized vocabulary: Standardized vocabularies are sets of agreed or standardized terms. Each standardized vocabulary is maintained and published by some authoritative body. Examples are names of countries and languages, classifications (e.g. food groups), units, methods, etc. Authoritative bodies may be ISO, CODEX, INFOODS, EUROFOODS, etc.

3. RECOMMENDATIONS

Recommendations 1 describes the complete data schema, i.e. all possible attributes and their domains, the relationships between the entity-sets and all additional entity-sets needed for implementation. It also defines which attributes are considered mandatory within the EUROFOODS data exchange framework and which ones are optional or recommended as further metadata (also see the appendix). *Recommendations* 2 lists and describes all sets of standardized vocabularies (thesauri) to be used in food composition data interchange. Some of the thesauri were developed from scratch, others were adopted from various international bodies. *Recommendations* 3 specifies constraints on the file formats to be used for data interchange and also describes procedures for data compression. *Recommendations* 4 specifies constraints on the media to be used for data interchange.

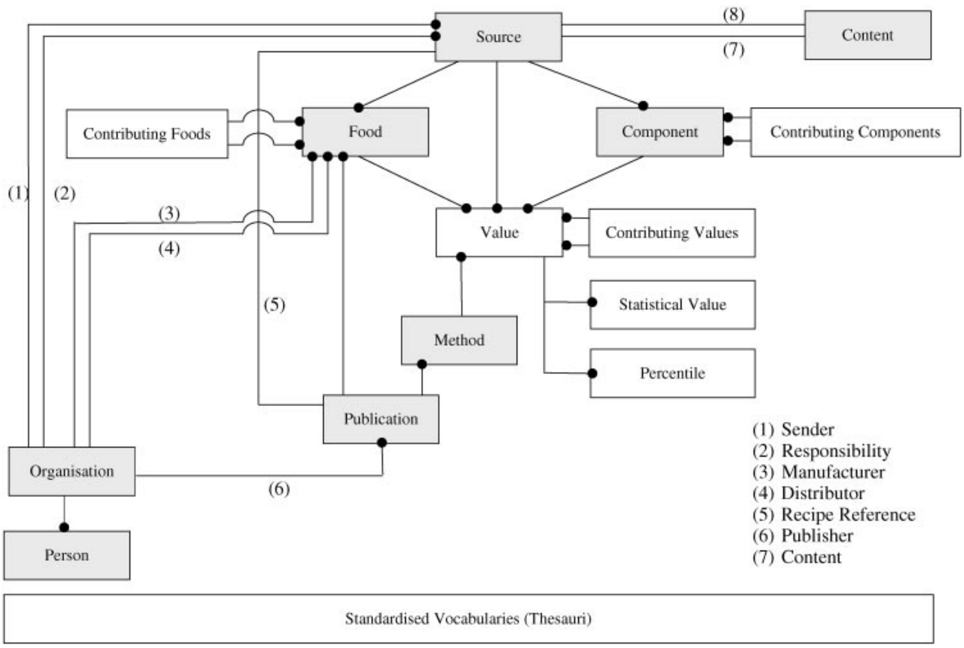

FIGURE 4. Complete entity relationship data schema: —● depicts a one-to-many relationship.

3.1. *Recommendations* 1: *Conceptual Data Schema*

Each food composition study to be interchanged (interchange package) is stored as a relational database. The relational approach was chosen because of the popularity of relational database systems, which allow for rapid implementation. A translation into object-oriented data models or into an XML document-type definition (DTD) (Connolly, 1997; WWW, 1999) would result in a more elegant representation of the data. This is possible and left to local data managers. The entity relationship schema depicted in Figure 4 is a refinement of the schema presented in Figure 3. The additional entity sets are necessary to store metadata to further describe *source*, *food*, *component*, *value* and *method* entities. The highlighted entity sets are implemented in a special way as described in Section 3.1.1. All relationships between entity sets are conditional, i.e. an entity in one set (in one table) does not necessarily have to be related to an entity in the related set. At the attribute level, however, we will classify some attributes as mandatory in order to guarantee a certain level of documentation in EUROFOODS interchange packages. A detailed list of attributes and their definition is given in the appendix.

3.1.1. *Special modelling and implementation of meta-tables*. The following requirements apply to the entity sets highlighted in Figure 4: *source*, *content*, *food*, *component*, *method*, *publication*, *organization* and *person*. We call these entity sets *meta-tables*. It should be possible to

- add further attributes in the future without much extra programming,
- interchange only those attribute values within a table that are actually used,

TABLE 1

Table schema of meta-tables

Attribute name	Short name	Data type[1]	Description
EntityID	ENTITYID	NUM	A unique number identifying the food, component, method, etc. See Section 3.1.2 for further information (Generic IDs)
PropertyID	PROPID	STR8	Max. 8 character property identification
Value	VALUE	STR255	The property value in text format. Properties of type MEM must be stored in the MemoValue field
MemoValue	MEMO	MEM	The property value in text format. Should be used for values of type *memo* (memo = longer than 255 characters) and for alternative free text values of properties of type *thesaurus*
Language	LANG	STR5	According to ISO 639 (1988a), a 2-character standard ISO language code plus an optional 2-character standard ISO country code separated by a blank character, e.g., "en" for English or "en UK" for British English
Preferred	PREF	BLN	True (1) indicates preferred terms, false (0) indicates synonyms. In case of blank values (NULL), True is considered the default value
Remarks	REMARKS	MEM	Free text annotations of the value

[1] According to Table 9.

- use set-valued attributes, i.e. attributes that hold more than one value,
- use several languages (translations) for textual data description,
- indicate preferred terms and multiple synonyms for a textual description,
- allow for free text- and thesaurus-based descriptions in parallel,
- annotate every single value if necessary,
- process the data with standard relational database management systems.

To meet these requirements using relational database technology, the corresponding entity sets are implemented using the schema given in Table 1. This technique allows description of an entity (a food, component, method, etc.) with an arbitrary number of property/value pairs in multiple languages, with multiple synonyms, and to attach annotations to every single value if necessary. Within such a table, each combination of [EntityID, PropertyID, Value/Memo Value, Preferred, Language] must be unique. Thus, these attributes form the key of the table. Note that in the remainder of the paper, the term *attribute* is used for attributes in the sense of column headers in relational tables, whereas the term *property* is used for names of properties in the property/value pairs described above.

3.1.2. *Formal conventions.* This section defines some formal conventions used for the complete schema description in the appendix.

Generic IDs: A *generic ID* is assigned to each entity (e.g., food, component, method, etc.) in each of the entity sets in the schema. IDs are implemented as positive integer values. It is left to the receiver of an interchange package to resolve the IDs to whatever system he or she uses to store multiple interchange packages in an integrated archival or reference database. IDs must be unique, i.e. no two entities can have the same ID. Further, IDs must be consistent within an interchange package, i.e. references to other entity sets must point to existing entities and all entities must be reachable through the primary source.

Isa-type relationships: Even though *isa-type relationships* are not directly supported by the relational data model, they often occur in real-life applications. As an example consider bibliographic reference data. Books, reports or journal articles are data sources ("a book *is a* source"). Different attributes are needed to describe a publication, depending on its type (e.g. a book has an ISBN whereas a journal article does not). For data interchange through relational databases, we propose a straightforward approach: e.g. all data on all sources are stored in a single table *Source*. Each source is assigned a property *publication type*, which triggers other applicable properties.

Properties: For each entity set in the database schema, a list of all possible properties is provided. Each property is given a name, a unique 8-character property-id to be used in interchange packages, a data type (see Table 9) and a priority. Further notes and explanations are provided for each property under scope note. The list of properties also shows the isa-type relationships in hierarchical form. Some properties are grouped for ease of discussion. Group headers are printed in italics and might be of interest for implementation in future editing or browsing software. As a default rule, a property *Remark* of type memo Memo (MEM) is assigned to each table within the database schema. This allows storing of all additional information not covered elsewhere in the schema.

Set-valued attributes: Data types of properties that allow more than one value are printed in brackets: e.g. {THS}. Data of type String (STR) and Memo (MEM) are always considered to be set valued in order to allow multiple translations of the text.

Priorities: The working group agreed that priorities of properties should be based on the level of operation. The lower the level, according to the four-level structure presented in Section 2.1, the more metadata are expected because the data reported are closer to their original source. The priorities given in the following sections should be interpreted as seen from a food composition data compiler's point of view. There are three priorities:

- *Mandatory* (M) properties build the core set of data that is needed to be able to capture the basic idea of a given food composition study.
- *Recommended* (R) properties should be considered the goal for everyone participating in data interchange.
- *Optional* (O) properties only apply to special circumstances and serve as a guideline to possible points of important data.

Priorities are also given for whole entity sets (i.e. tables). If a recommended or optional entity set is used, the priorities for its properties apply as indicated in that entity set.

Complementary use of thesaurus-based values and free text: Properties that use THS as their data type only allow values that are part of the corresponding thesaurus. If for some reason the given thesaurus is not adequate, if a certain term is missing in the thesaurus, or if free text description is preferred over standardized vocabulary, the MEMO attribute should be used instead of the VALUE attribute. Further remarks should be placed in the REMARKS attribute. This mechanism allows the user to use both systems in parallel or to introduce new terms that might become standard in the thesaurus at a later time.

3.2. *Recommendations 2: Standardized Vocabulary*

This section lists those standard vocabularies (thesauri) that have been especially developed for the recommendations. Other thesauri are adopted from existing

TABLE 2

Table schema for standardized vocabulary

Concept property	Description
Code	A unique and short alphanumeric code identifying each concept. The code is mainly used in data interchange packages and does not necessarily need to be self-explanatory. Codes are *not* case-sensitive. Codes are kept unchanged when translating a thesaurus
Descriptor	A text-string describing the concept. This string, like the code, must be unique since it is the representation of the code to the user
Scope note (optional)	A longer text explaining in detail any specialities to be considered when applying the concept (e.g. exceptions, relation to other concepts, further clarifications and definitions)
Synonyms (optional)	Synonymous text strings that express exactly the same concept as the descriptor and help people to find a concept (e.g. vitamin B1 and thiamin)
Abbreviation (optional)	Like the descriptor, but limited to 32 characters for computer processing with limited screen space

standards. For references see the appendix. Each thesaurus consists of a set of concepts that may be arranged within a hierarchy. A concept is represented by a main descriptor—a term representing the concept—and may be further described with a scope note and synonyms. The official thesauri will use English as their main language. It is up to each user to translate thesauri for local usage. However, it is recommended to establish a central authority within each country to maintain and publish translations. It is also a wise idea to share translations among countries using the same language (e.g. Germany, Austria and Switzerland). EUROFOODS will try to keep track of existing translations. This information will be accessible through the EUROFOODS homepage. The fields describing the concepts within a thesaurus are given in Table 2. Further fields for version control of concepts are available within the Thesaurus Manager software.

3.2.1. *Publication types.* The publication type lists general terms for describing ways of publishing food composition data (Table 4).

3.2.2. *Acquisition Types.* The acquisition type lists general terms for describing general categories of food composition data sources and methods of data acquisition, e.g. laboratory, food composition compiler, food industry claim, etc. (Table 3).

3.2.3. *Value types.* The value type is designed to further describe the figure in *Best Location* in the *Value* table, or to give a qualitative description of the value when no *Best Location* can be given (Table 7).

3.2.4. *Units.* Unit description is influenced by International Standard (ISO 1000, 1992a). The standard is extended with food composition specific units. Table 5 lists the units that have so far been identified to be relevant to the field.

3.2.5. *Modes of expression.* See Table 7.

3.2.6. *Method types.* The method type lists general terms to describe how a value was obtained or generated (Table 6).

TABLE 3

Acquisition types

Code	Descriptor	Scope note
O	In-house or affiliated laboratory	(O = own); in-house or affiliated laboratory report/protocol. Study design, sampling, and analysis are under direct control of the person or organization reporting the data
I	Industry laboratory	Laboratory report/protocol of a food producer or distributor
D	Independent laboratory	Laboratory report/protocol of a third-party laboratory not directly affiliated to the food producer or the organization that initiated the investigation and now reports the data
F	Food composition table	Compiled food composition table. The compiler is now responsible for the data. Typically, the underlying data sources are only documented briefly but further information is available from the compiler. Food composition tables are mostly published by the compiler
P	Published and peer-reviewed scientific paper	Peer-reviewed study, published in a journal or book
L	Food label, product information	Food label or product information provided by the producer or distributor with no further information about the data sources
S	Value created within host-system	To be used for values created by a compiler within his or her FCDMS using calculation or estimation. Note: simple unit conversion does not fall into this category
E	Other acquisition type	(E = else); other acquisition type not mentioned in this list
X	Acquisition type not known	

TABLE 4

Publication types

Code	Descriptor	Scope note
B	Book	
AB	Article in book	
J	Journal	
AJ	Article in journal	
R	Report	
AR	Article in report	
AD	Authoritative document	Document published by legal authorities, standards organizations, committees, patent offices, etc.
F	File or database	
SW	Software	
L	Product label	
P	Personal communication	Personal communication with no further bibliographic information but the reporter's name and address
E	Other publication type	(E = else); other publication type not mentioned in this list
X	Publication type not known	

3.3. *Recommendations* 3: *File Formats for Data Interchange*

Recommendation 3 covers technical aspects of data interchange. It describes how the tables of a relational database structured according to recommendation 1 should be formatted for transmission on disk or via the Internet.

TABLE 5

Units

Code	Descriptor	Scope note
RE	Retinol equivalent	1 RE = 1 μg all-*trans* retinol
BCE	beta-Carotene equivalent	1 BCE = 1 μg all-*trans* beta-carotene
ATE	alpha-Tocopherol equivalent	1 ATE = 1 mg RRR-alpha-tocopherol 1 ATE = 1 mg d-alpha-tocopherol
NE	Niacin equivalent	1 NE = 1 mg niacin or 60 mg tryptophan
MSE	Monosaccharide equivalent	1 MSE = 1 g glucose
kg	kilograms	
g	grams	
mg	milligrams	
ug	micrograms	
ng	nanograms	
L	litres	
mL	millilitres	
uL	microlitres	
mmol	millimols	
kJ	kilojoules	
kcal	kilocalories	
R	Ratio	

TABLE 6

Modes of expression

Code	Descriptor	Scope note
W	per 100 g edible portion	
T	per 100 g total food	As purchased including any waste, e.g. chickenwing with bones, banana including peel, etc.
D	per 100 g dry weight	
V	per 100 mL food volume	
WKG	per kg edible portion	
TKG	per kg total food	
DKG	per kg dry weight	
VL	per L food volume	
F	per 100 g total fatty acids	
N	per g nitrogen	

3.3.1. *Text encoding*. All data must be transmitted in textual form in order to be interpreted on the widest range of computer platforms possible. Text must be encoded using either ISO/IEC 646 (1991) (7-bit code) or ISO/IEC 8859-1 (1998e)/(8-bit code). The use of Unicode is planned for future implementations, when this system is fully established on the market.

3.3.2. *File format*. Each database table must be stored in one text-file with one data record per line. The names of the tables are listed in Section 3.3.6. Tables that are not used can be omitted.

Data fields should be delimited by semicolons (; =ASCII 59). The delimited file format has two advantages compared to fixed-length record files: reduced file size and easy handling of memo-fields (i.e. text fields larger than 255 characters). Text and

TABLE 7

Value types

Code	Descriptor	Scope note
MN	Mean	The compiler chose the mean of the statistic as *Best Location*
MD	Median	The compiler chose the median of the statistic as *Best Location*
MI	Minimum	The compiler chose the minimum value within the statistic as *Best Location*
MX	Maximum	The compiler chose the maximum value within the statistic as *Best Location*
W	Weighted	The *Best Location* is a weighted average of values from several sources. Examples of weighting criteria include weighting by brands, weighting by number of samples, etc.
LT	Less than	Use this value type if there is no further statistical information available for MX and if no other value type applies. LT is also useful in case of calculated or imputed rather than analysed values. The figure given in *Best Location* should be interpreted as an upper limit
MT	More than	Use this value type if there is no further statistical information available for MN and if no other value type applies. MT is also useful in case of calculated or imputed rather than analysed values, e.g. in recipe calculation. The figure given in *Best Location* should be interpreted as a lower limit
BE	Best estimate	According to the responsible compiler, the value is the "best" available. This type should be used when there is no further statistical information available
TR	Trace	Use Trace only when there is evidence that some amount of the component is present but no precise figure can be given, e.g. if the level measured is below the level of quantification. Further information about the exact definition of trace should be provided under *Remarks* in either the corresponding Value-, Method-, Component-, or Source-Description. Normally trace values have a blank *Best Location.* Never use trace together with a zero in *Best Location*
BL	Below detection limit	The component is not detectable with the applied method, e.g. below the limit of detection. However, the component might be present. It is recommended to provide information about the limit of detection within the corresponding method description. Use BL together with a blank *Best Location*
LZ	Logical zero	The component in question never appears in the food in question, e.g. alcohol in meat, or fat in mineral water. Use LZ together with Method Type E
RZ	Regulatory zero	The component in question never appears in the food in question according to (national) food regulations
UD	Undecidable	Use this value type together with a blank *Best Location* in cases where no decision can be made, e.g. the available data differ too much. Other statistical parameters, however, might be available, e.g. minimum and maximum
N	Unknown	Use this value type together with a blank *Best Location* in cases where compilation work has shown the value to be unknown, i.e. there is no literature available and no estimation or calculation possible. This Value Type is useful in food composition tables and might be useful at other levels of the compilation process (see Figure 1)
E	Other value type	(E = else); other method type not mentioned in this list
X	Value type not known	The type for the given value is not known

memo fields must be enclosed in double quotes (" = ASCII 34). Alternatively, the fixed-length file format may be used to support a wider range of software on the various computer platforms. Another advantage is better legibility if the file is viewed in a text editor. Memo fields, however, may vary and a maximum length must be computed for each field in advance. In both cases, the first line in the file must contain the standardized field names as given in recommendation 1. In case of fixed-length files, the field name must be followed with its length in brackets (see Example 2).

TABLE 8

Method types

Code	Descriptor	Scope note
AG	analytical, generic	Use this Method Type if no further information on the nature of analysis is available
A	analytical result(s)	Analytical result or statistic of multiple measurements of the same sample (replicates). See the property *Headline method name* in the *Method* table for further information
D	aggregation of contributing analytical results	Value derived as an aggregation of accepted analytical contributing results (e.g., from different sources). See the property *Headline method name* in the *Method* table for further information
CG	calculated, generic	Use this Method Type if no further information on the nature of calculation is available
G	calculated as aggregate food item	Used in case of aggregated foods when the composition is mainly obtained by summation of the composition of its ingredients. See food description for further information
R	calculated as recipe	Used in case of complete recipe calculation including NLG factors. See food description for further information
P	calculated on component profile	For example, fatty acid profile, amino acid profile for a specified food. See component description for further information
S	summation from constituent	See component description for further information. Note that summation includes subtraction, e.g. calculation of total carbohydrates by difference
T	calculations including conversion factors	For example, for energy calculation or for calculating alpha-tocopherol equivalents. The conversion factors should be documented within the recursive value description or within the method or component description
K	calculated from related food	Useful as a separate case where a specific calculation, rather than imputation is performed on a related food, e.g. toast from bread or calculating the values for a food "weighed with waste". The food description should link to the related food
IG	imputed/estimated, generic	Use this Method Type if no further information on the nature of imputation/estimation is available
I	imputed/estimated from related food	The food description should link to the related food. No further information on the method is available
O	imputed/estimated from other foods and other related components	Note that with *food* and *component* we refer to the definitions given in Section 2.2
L	estimated according to regulatory requirements	L stands for legislation
E	Other method type	(E = else); other method type not mentioned in this list
X	Method type not known	No method information is available

Example 1 (;-*delimited*):

FRSTNAME; LASTNAME; EMAIL
"Anders"; "Møller"; "amoeller@vfd.dk"
"Wulf"; "Becker"; "wulf.becker@slv.se"
. . .

Example 2 (*fixed-length*):

FRSTNAME(15)	LASTNAME(15)	EMAIL (18)
Anders	Møller	amoeller@vfd.dk
Wulf	Becker	wulf.becker@slv.se

. . .

TABLE 9

Data types for data interchange

Data type	Textual representation	Example
STR*nnn*	Text String with a maximum of *nnn* characters where *nnn* stands for a number between 0 and 255. 255 applies if no length is specified. Double quotes (“ = ASCII 34) are not allowed in strings. Use single quotes (‘ = ASCII 39) instead	Jayne Ireland
MEM	Memo: text strings larger than 255 characters. Double quotes (“ = ASCII 34) are not allowed in strings. Use single quotes (‘ = ASCII 39) instead	A verbose comment, lengthy explanations etc
DAT	Date: generally in the form CCYY-MM-DD with leading zeros (ISO 8601, 1988b). In case of reduced precision, days (DD), months (MM) or years (YY) may be omitted starting from the extreme right-hand side. If time is also relevant use CCYY-MM-DD/hh:mm:ss	1999-01-21 1999-07 1984 1997-12-03/21:35:01
INT	Integer: in the range of ± 2147483648 ($= \pm 2^{31}$)	165
NUM	Decimal numbers: use the point (. = ASCII 46) to separate decimals. All given decimals must be significant. Do not cut trailing zeros, i.e., trailing zeros should be used to indicate significant decimals	3.472 5.0
FRC	Fraction: a decimal number between 0 and 1 (0 and 1 inclusive)	0.34
BLN	Boolean: 1 = true, 0 = false	0
THS	Thesaurus entry: use valid interchange codes in string format	B0123
FIL	Additional (multimedia) Files: Generally, files are referred to as URLs. If a leading “http://” or “ftp://” is omitted, “file://MMFILES/” is the default, i.e. a simple filename refers to a file in the directory MMFILES which is part of the interchange package. Files must use 8-character long filenames with an up to 3-character long file extension (also see Table 10). Future versions of the recommendations will allow for longer filenames	IMG123.JPG http://xyz.com/images/apple.gif ftp://abc.org/docs/manual.doc
KEY/FKY	Keys and foreign keys: Positive integers > 0 as described in Section 3.1.2	136523

3.3.3. *Data-type formats.* Within an interchange package, the data types given in the database schema in recommendation 1 must use the text formats given in Table 9.

3.3.4. “*Readme*” *file.* Extra information extending the recommendations (e.g. further text or database documents) may be added and must be described in a text file (README.TXT) using text encoding according to Section 3.3.1. The file format specifications concerning field separation of the database tables must be specified within the README.TXT file.

3.3.5. *Bundling and compression of files.* For ease of handling and to reduce data size, the whole database as described in recommendation 1 can be compressed and bundled into one file. The following rules apply for file compression: It is recommended to use ZIP-compression. The ZIP format is widely used and software for decoding is available on many platforms. Within a compressed archive paths relative to the root directory should be used. Self-extracting archives (.exe) that can be run under the MS DOS operating system should be used only with bilateral agreement.

3.3.6. *Directory structure and filenames.* The files that form a food composition database should be named and arranged as given in Table 10. All files within the “DB”-directory must be present even if they do not contain any data.

TABLE 10

Directory structure for interchange packages

File/Directory name	Explanation
EFXvvaaaa/	The whole interchange package, i.e. all files, should be stored in one directory. We suggest naming such a directory according to the schema given on the left. "EFX" stands for EUROFOODS File Exchange. "vv" denotes the version number of the interchange recommendations used. The remaining characters can be chosen arbitrarily to distinguish separate packages. Example: EFX10ab4
DB/	Directory "DB" contains all database files
SOURCE.TXT	Table Source
CONTENT.TXT	Table Content
FOOD.TXT	Table Food
CONTFOOD.TXT	Table Contributing Food
COMPONEN.TXT	Table Component
CONTCOMP.TXT	Table Contributing Component
VALUE.TXT	Table Value
CONTVAL.TXT	Table Contributing Value
STATVAL.TXT	Table Statistical Values
PERCENT.TXT	Table Percentiles
METHOD.TXT	Table Method
PUBLICAT.TXT	Table Publication
ORGANISA.TXT	Table Organization
PERSON.TXT	Table Person
MMFILES/	Directory "MMFILES" contains all multimedia files mentioned in the database. Basically, every file type is allowed (e.g. Word.doc, Acrobat.pdf, Rich Text Format.rtf, ASCII-Text.txt, Access.mdb, Excel.xls, dBASE.dbf, etc.). However, preference should be given to the most widely used file types. For image files, preference should be given to JPG-files (.jpg) or eventually GIF-files (.gif). These file formats use data compression (unlike TIFF-files)
README.TXT	The "readme" file (see Section 3.3.4)

3.4. *Recommendation* 4: *Media to Use for Data Interchange*

Food composition data packages as described in the previous recommendations should be exchanged using either a physical storage device or the Internet as a transportation medium. The following basic rules should be applied to guarantee maximal system compatibility on the physical level.

3.4.1. *Physical storage devices.* Only diskettes and CD-ROMs should be used for data interchange. In case of doubt about the technical facilities of the receiver, diskettes should be preferred.

Diskette: DOS-formatted PC-diskettes with 1.44 MB capacity should be used since Macintosh and Unix systems can handle this format too.

CD-ROM: No further restrictions apply to CD-ROMS since all CD-ROM systems adhere to the same international standard (ISO 9660; 1988f). Note that there is a trend towards DVD (Digital Versatile Disc). DVDs will be recommended as soon as this standard is established and widely available on the market.

3.4.2. *Internet.* If data files are transferred over the Internet using e-mail, File Transfer Protocol (FTP) or the World Wide Web (WWW), the following rules should be applied:

E-mail: The names and formats of all attached files should be mentioned in the e-mail body. Files should be sent as MIME compliant e-mail attachments (MIME = Multipurpose Internet Mail Extensions see Internet RFC 2045, 2046 and 2049). Proprietary solutions only available within certain mailing tools, intranets or computer platforms should be avoided.

FTP: FTP allows the transfer of files in text mode or binary mode. Binary mode should be used in all cases to preserve the original file structure and prevent the conversion of text into proprietary representations.

WWW: The authors are not aware of specific problems concerning the transmission of data files via WWW using the HTTP protocol.

4. TOOLS FOR IMPLEMENTATION

To promote the implementation of the EUROFOODS recommendations, we are developing software tools for various tasks. Since this development is an ongoing process, we will only give a brief summary of the current status. Further results will be presented via the EUROFOODS homepage at http://food.ethz.ch/cost99/. The following tools are available or under development.

- *Template for data schema*: the data schema as described in recommendation 1 is available as a Microsoft Access file.
- *Template for standardized vocabulary database*: This Microsoft Access file allows the user to build a repository of multiple standardized vocabularies (thesauri).
- *Thesaurus Manager*. This Microsoft Windows-based software tool can be used to build, maintain and translate consistent standardized vocabularies (i.e. thesauri, mono- and poly-hierarchic classifications, grouping systems, lists of terms, etc.). It is planned to modify this tool to access the Microsoft-Access-based standardized vocabulary database via ODBC. Currently, this tool stores a standardized vocabulary in dBASE IV format.
- *Repository for multiple interchange packages*. This Microsoft Access file extends the data schema for single interchange packages in order to store multiple interchange packages. A Microsoft-Access-based program code is provided to import interchange packages which follow the recommendations.
- *Generic WWW interface to the repository*: This software is based on Microsoft's active server pages (ASP) and allows the user to browse the repository for multiple interchange packages over the internet. It also uses the standardized vocabulary database described above.
- *Processing of complex queries*: A WWW-based application is available via the LanguaL homepage (http://food.ethz.ch/langual/) to demonstrate the possibilities of thesaurus-based database queries. The foods of four national food composition databases (France, Denmark, Hungary, U.S.A.) can be searched in a flexible way using LanguaL thesauri (Hendricks, 1992).
- *Data Editor/Browser*: A Microsoft Excel based editor and browser is under development. This tool helps the user to build and to view interchange packages according to the recommendations with a convenient Windows-based graphical user interface. Since this tool is based on Excel, import and export of data from other software is straightforward.

5. DISCUSSION AND FUTURE DIRECTIONS

The present recommendations for food composition database management and data interchange have been designed to be implemented using relational databases. Thus, we are able to build on existing and widely used technology. The relational database approach, however, has some disadvantages as discussed in Section 3.1. This problem is well known in the field of scientific and statistical databases (Shoshani, 1991).

Furthermore, to interchange a complex relational database, several files must be interchanged (one for each table). It is therefore planned to translate the data schema presented in this paper into an XML application (extensible markup language) (Connolly, 1988; WWW, 1999) once this Internet standard has been established. All data within an interchange package could then be stored in a single text file following a well-defined grammar. XML offers conceptual and technical solutions for the problems mentioned above because data can be treated in a more object-oriented way. XML is a meta-language for the design of markup languages such as HTML. A markup language defines the way to describe information in a certain class of documents (e.g. HTML and hypertext). In contrast, XML allows the user to define customized markup languages for many classes of documents. XML is a simplified dialect of SGML (Herwijnen, 1994) and was designed to make it easier to use SGML on the WWW.

Using XML, we would be able to define a food data markup language (FDML). It would share the idea of marked-up text as proposed by the INFOODS data interchange format (Klensin, 1992). Our design described in Section 3.1, however, would lead to a different structure in order to overcome the drawbacks described by Unwin and Møller (1996).

Another area that needs further development is recipe management and interchange. Although it is possible to handle recipes (i.e. ingredients and recipe procedure) in the current recommendations, it is not yet possible to interchange important information used in recipe calculation such as nutrient loss and gain factors and weight loss factors (yield).

Although the recommendations have been successfully implemented in Switzerland and form the technical basis for the Swiss food composition program, further multinational trial studies are needed to investigate the usefulness of the proposed properties within the metadata schema (recommendations 1 and 2).

6. CONCLUSIONS

We developed recommendations to enable consistent *data interchange* between food composition data producers, compilers and users. The recommendations, however, do not explicitly interfere with, or cover, internal laboratory management procedures. The recommendations are especially useful to disseminate national food composition tables or to interchange specific food composition studies among data compilers.

The first part of the recommendations defines the kinds of descriptive data to be considered when dealing with food composition data. It also suggests standardized vocabulary to be used in these descriptions. The second part covers technical issues of data interchange, such as file formats and media to use for data transfer.

The recommendations can be implemented at various levels of detail and allow for future extensions. In particular the properties used to describe foods, components, compositional values, analytical and other methods, and data sources, together with their standardized vocabulary, can be extended without affecting the core design of the recommendations. Thus, it is not necessary to rewrite software tools after such changes. Using this framework people can gradually extend schemes of metadata and vocabularies. The market and future experience will decide what types of metadata will form the core for future recommendations for food composition data interchange world-wide.

Besides allowing for data interchange, the recommendations can also be used as a framework to develop *data management* procedures for archival and reference databases. The Swiss team implemented a prototype archival database (repository) to store multiple interchange packages and allow for uniform data access across such data sources. Data access is available through a WWW interface or using Microsoft Excel spreadsheet applications. The development of software tools to facilitate the implementation of the recommendations is an ongoing project. Further information is available on the internet via http://food.ethz.ch/cost99/datax/.

The authors would like to thank the EU COST Action 99, “research action on food consumption and composition data”, for the financial support of various meetings that made this work possible.

APPENDIX: DATABASE TABLE DEFINITIONS

TABLE A11

Source table

Property name	Property ID	Data type	Prio (prim.)	Prio (sec.)	Scope note
Source name	SRCENAME	{STR}	R	R	The Source Name should be kept short and should include important parts of the responsible organization and/or the bibliographic reference
Primary source	PRIMSRCE	BLN	M	n.a.	True (1) if the data source is the primary source within the interchange package
Compilation	COMPLANG	THS	M	M	Language, originally used for free text description within all data of a source (incl. food, component and value description). According to ISO 639: a 2-character standard ISO language code plus an optional 2-character standard ISO country code separated by a blank character, e.g. “en” for English or “en UK” for British English
Acquisition type	ACQTYPE	THS	M	M	According to Table 3
Responsibility	RSPONSIB	FKY	M	M	Link to the organization table (and thereby to the person). ID of the organization that is responsible for the content of the data source
Sender	SENDER	FKY	R	O	Link to the organization table (and thereby to the person). The ID of the organization that sent the interchange package
Sent date	SENTDATE	DAT	M	M	The date the interchange package is sent
Legal restrictions	LEGLREST	{MEM}	R	R	Note any legal (copyright) or scientific restrictions imposed on the data. Such information is also known as *disclaimer*
Content summary	CONTSUMM	FKY	R	R	Link to the content table (Content ID). Briefly describes the content of an interchange package
Excluded content summary	EXCONSUM	FKY	O	O	Link to the content table (Content ID). Briefly describes what data have been omitted compared to the original data source. Use this attribute when an interchange package represents just a part of a more comprehensive data source. This information might help people to localise further data
Bibliographic refernce	BIBREF	FKY	M	M	Link to publication table (Publication ID)
Original food groups	ORIGFDGP	FIL	R	O	A file listing the original food groups and their codes. Preference should be given to a plain text file. There is currently no further specification on the format of this file
Quality assessment	QUALASSM	FIL	R	O	Link to a file describing the meaning of quality indices, scores, criteria used, expert systems used, etc. for the assessment and documentation of the quality of each compositional value (see attribute QI in value description)
Remarks	REMARKS	{MEM}	O	O	Any further remarks

TABLE A12

Organization table

Property name	Property ID	Data type	Prio	Scope note
Organization name	ORGNAM	{STR}	M	The official name of the organization
Super organization name	SPORGNAM	{STR}	O	If applicable, give the name of the umbrella organization
Postal address	POSTADDR	{MEM}	R	Postal address as would be put on a letter, i.e. PO box, address, ZIP-code, city, country, etc.
Country	COUNTRY	THS	M	Use ISO 3166-1 (1997). A country subdivision code as described in ISO 3166-2 (1998b) can be added after the country code separated by a hyphen, e.g., CH-ZH.
Telephone	PHONE	{STR}	R	Telephone and fax numbers should be formatted from an international point of view. Use the form + country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen
Fax	FAX	{STR}	R	Should be formatted from an international point of view. Use the form + country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen
E-mail	EMAIL	{STR}	R	Internet e-mail address
WWW	WWW	{STR}	R	Always give complete URLs. Example: http://www.fao.org/infoods/
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A13

Person table

Property name	Property ID	Data type	Prio	Scope note
Organization	ORGID	FKY	M	Link to organization table. Gives the ID of the organization to which the person is affiliated
Title	TITLE	{STR}	R	The title used to address a person, e.g. Prof or Dr. If there is no title or in case of doubt, use Mr or Mrs
First names	FRSTNAME	{STR}	R	Separate multiple names with space characters Abbreviations are allowed
Last name	LASTNAME	{STR}	M	Family name of the person
Position	POSITION	{STR}	R	The current working position of the person, e.g. laboratory director, nutritionist, IT manager, etc.
Postal address	POSTADDR	{MEM}	R	Complete postal address as would be put on a letter
Country	COUNTRY	THS	M	Use ISO 3166-1. A country subdivision code as described in ISO 3166-2 can be added after the country code separated by a hyphen, e.g. CH-ZH
Telephone	PHONE	{STR}	R	Should be formatted from an international point of view. Use the form + country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen
Fax	FAX	{STR}	R	Should be formatted from an international point of view. Use the form + country-code area-code sub area-code phone-number. The various blocks should be separated with a space character or hyphen
E-mail	EMAIL	{STR}	R	Internet e-mail address
WWW	WWW	{STR}	R	Always give complete URLs. Example: http://www.fao.org/infoods/
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A14

Content table

Property name	Property ID	Data type	Prio	Scope note
Food description	FOODDESC	{MEM}	O	Free text describing what techniques are used to describe foods
Number of foods	NRFOODS	INT	M	The total number of foods in the data source
Food types				
Basic foods	BASICFDS	FRC	R	Percentage of basic or generic raw and processed foods in the data source, e.g. meat, fish, fruits, vegetables, and products
Brand named food products	BDFDP-RODS	FRC	R	Percentage of raw or processed foods of specific brands
Dishes	DISHES	FRC	R	Percentage of dishes, i.e. meals and recipes that can be produced in home kitchens using basic foods and food products
Main food groups	FOOD GRPS	{THS}	R	List the food groups of the foods in the interchange package. Use principal food groups of Eurocode 2 (Kohlmeier, 1992)
Component description	COMPDESC	{MEM}	O	Free text describing what techniques are used to describe components
Number of components	NRCOMPS	INT	M	The total number of components in the data source
Component groups	COMPGRPS	{THS}	R	List component groups covered by the data source
Value description	VALDESC	{MEM}	O	Free text describing what techniques are used to describe values
Value sources				
Own analysis	OWNANALY	FRC	R	Percentage of values obtained by own analysis, i.e. all data that have been analysed by the data compiler's own or affiliated lab
Other analysis	FORNALY	FRC	R	Percentage of values by other analysis, i.e. using other analysis, for all data sources that were not produced under the compiler's initiative or supervision, i.e. data someone else published before
Calculation	CALCUL	FRC	R	Percentage of values obtained by calculation
Estimation	ESTIMAT	FRC	R	Percentage of values obtained by estimation
General use	GENRLUSE	{MEM}	O	Free text description of the data's target user group and scientific restrictions. It might also be useful to indicate countries or regions where the data are applicable or not
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A15

Publication table

Property name	Property ID	Data type	Prio	Scope note
Title	TITLE	{STR}	M	The title of the publication. Use this property several times to provide the title in the original language, in English, and any other language if possible
Authors	AUTHORS	{STR}	M	Separate all multiple authors by semi-colon (;). For personal names, write the forename or initials after the last name, separated by comma. The attribute may be

TABLE A15 (*Continued*)

Property name	Property ID	Data type	Prio	Scope note
				used for the name of an organization where this is considered a corporate author, for example "AOAC", or for the abbreviation "Anon." where the authorship is anonymous
Publisher	PUBLISHR	FKY	M	Link to the organization table. The ID of the organization that published the publication
Publication date	PUBDATE	DAT	M	The year or exact date, the publication was issued
Version	VERSION	{STR}	O	Use this attribute for any versioning system other than publication date or edition number. This attribute is helpful for frequent updates
Original language	ORIGLANG	THS	M	The language that the publication was originally written in. According to ISO 639: a 2-character standard ISO language code plus an optional 2-character standard ISO country code separated by a blank character, e.g., "en" for English or "en UK" for British English
Languages	LANGS	{THS}	R	Language codes of all other languages, that major parts of the publication have been translated into. According to ISO 639: a 2-character standard ISO language code plus an optional 2-character standard ISO country code separated by a blank character, e.g. "en" for English or "en UK" for British English.
Publication type	PUBTYPE	THS	M	The publication type triggers further metadata (see below). According to Table 4
If a book				
ISBN	ISBN	{STR}	R	International Standard Book Number
First edition date	FSTEDAT	DAT	O	When was the first edition published?
Edition number	EDNR	INT	R	What is the current edition?
Number of pages	NRPAGES	{STR}	O	Total number of pages
If an article in book				
Book title	BKTITLE	{STR}	M	The title of the book in which the article appears. The title of the article is given in the TITLE property
Editors	EDITORS	{STR}	M	The names of the editors of the book
ISBN	ISBN	{STR}	R	International Standard Book Number of the book
Pages	PAGES	{STR}	O	The book pages covered by the article, e.g. 45–47
If a journal issue				
Long journal name	LGJRNAME	{STR}	O	
Abbreviated journal name	ABJRNAME	{STR}	M	
ISSN	ISSN	{STR}	O	
Volume	VOLUME	{STR}	M	
Issue	ISSUE	{STR}	M	
If a journal article				
Long journal name	LGJRNAME	{STR}	O	
Abbreviated journal name	ABJRNAME	{STR}	M	
ISSN	ISSN	{STR}	M	
Pages	PAGES	{STR}	R	The pages covered by the article, e.g. 375–383

TABLE A15 (*Continued*)

Property name	Property ID	Data type	Prio	Scope note
Volume	VOLUME	{STR}	M	
Issue	ISSUE	{STR}	M	
If a report				
Series name	SERINAME	{STR}	O	Use this property if the report is published within a series of other reports
Series number	SERINR	{STR}	O	The number of the report within the series
ISSN	ISSN	{STR}	M	
ISBN	ISBN	{STR}	O	
If an article in report				
Editors	EDITORS	{STR}	M	The names of the editors of the report
Report title	RPRTITLE	{STR}	M	The title of the report. The title of the article is given in the TITLE property
Series name	SERINAME	{STR}	O	
Series number	SERINR	{STR}	O	
ISSN	ISSN	{STR}	M	
ISBN	ISBN	{STR}	O	
Pages	PAGES	{STR}	R	The pages of the report covered by the article, e.g. 45–67
If a file or database				
File format	FILEFRMT	{STR}	M	Give information about the platform or computer system which the file is compatible to. Also mention the software needed to interpret the file
WWW	WWW	{STR}	O	The internet address (URL) of the file (WWW or FTP)
Publication medium	MEDIUM	{STR}	R	How is the file distributed: e.g. diskette, CD-ROM, tape, internet, etc.
If a software				
Operating system	OS	{STR}	M	Under which operating system (including version number) does the software run?
Primary publication media	MEDIA	{STR}	R	On what media is the software published, e.g. CD-ROM?

TABLE A16

Food table

Property name	Property ID	Data type	Prio	Scope note
Source	SOURCEID	FKY	M	Link to the data source reporting the food
Food name and identification				
Food name	FOODNAME	{STR}	M	The preferred food name and additional synonyms in various languages. Food names should start with an upper-case first character in the first word, e.g. Grapefruit, Spanish lime, etc. Scientific names must use Latin (la) as language flag and should adhere to the following format: Genus species Author [, Year] e.g. Gadus morhua Linnaeus, 1758
Abbreviated food name	ABBREV	{STR32}	O	Used for applications with limited screen/paper space
Original food code	ORIGFDCD	{STR}	R	The food code, ID or abbreviation used to identify the food in the original publication

TABLE A16 (*Continued*)

Property name	Property ID	Data type	Prio	Scope note
Original food group code	ORIGGPCD	{STR}	R	The proprietary classification code used in the original publication. The proprietary classification system should be provided separately under ORIGFDGP within the primary source description
Standard classification			M	At least one of the standard classification systems is mandatory
Product type	PRODTYPE	{THS}	R	FDA product type thesaurus of LanguaL facet A (Møller and Ireland, 2000)
CODEX food standards	CDXFDSTD	THS	O	Codex Alimentarius Food Standards code (Codex, 1989)
CODEX food categorization system for the general standards for food additives	CDXFDADD	{THS}	O	According to Codex (1996)
CODEX classification of foods and feeds	CDXFDFD	THS	O	According to Codex (1993a)
CODEX food categorization system for contaminants	CDXCONT	THS	O	According to Codex (1997a)
FAO food balance sheet classification	FAOFBS	THS	O	According to Trichopoulou and Lagiou (1997)
CIAA food categorization	CIAA	{THS}	R	According to CIAA (1994)
Eurocode 2	EC2	{THS}	R	According to Kohlmeier (1992)
European article number	EAN	{STR}	R	For European articles only
Universal product code	UPC	{STR}	R	
E-number	ENR	THS	R	If food is food additive, code according to the European E-number system for additive standardization
INS-code	INS	THS	R	If the food is a food additive, code according to the International Numbering System for food additives according to Codex Alimentarius
General description				
Manufacturer	MANUFACT	{FKY}	R	Link to organization table (Organization ID). Describes the direct manufacturer or producer of the food, e.g. a farmer is considered a manufacturer
Distributor	DISTRIB	{FKY}	R	Between producer and retailer. Link to organization table (Organization ID)
Food source	FOODSRCE	THS	R	Langual facet B (Møller and Ireland, 2000)
Genetically modified	GENMANIP	BLN	O	
Agricultural production conditions	AGRICOND	{MEM}	O	Brief description of soil conditions, watering schemes, feeding, harvesting, slaughtering, ripeness, etc.

TABLE A16 (*Continued*)

Property name	Property ID	Data type	Prio	Scope note
Colour	COLOR	{STR}	O	Colour values are currently not further specified. More detailed recommendations are planned for further versions
Generic image	GENIMAGE	{FIL}	R	The file names of generic images showing foods similar to the food or sample in question
Specific image	SPCIMAGE	{FIL}	R	The file names of specific images of the food sample, i.e. the food that was actually analysed
Part of plant or animal	PARTPLAN	THS	R	Langual facet C (Møller and Ireland, 2000)
Percentage edible portion	EDPORT	FRC	R	May also be considered a component
Nature of edible portion	NATEDPOR	{STR}	R	Which parts of the food are edible, e.g. flesh, root, leaf, etc.?
Nature of waste	NATWASTE	{STR}	R	Which parts of the food are not edible, e.g. rind, bone, stone, peel, etc.?
Physical state shape or form	PHYSTATE	{THS}	R	Langual facet E (Møller and Ireland, 2000)
Extent of heat treatment	HEATREAT	THS	R	Langual facet F (Møller and Ireland, 2000)
Treatment applied	TREATAPP	{THS}	R	Langual facet H (Møller and Ireland, 2000)
Cooking method	COOKMETH	{THS}	R	Langual facet G (Møller and Ireland, 2000)
Recipe procedure	RECPROC	{MEM}	R	If food is a recipe
Recipe bibliographic reference	RECREF	FKY	R	Link to publication table (Publication ID). Describes the publication holding the recipe
Final preparation	FINLPREP	{STR}	R	Final preparation of food before consumption, e.g. heating a frozen dinner or canned food
Preservation method	PRESMETH	{THS}	R	Langual facet J (Møller and Ireland, 2000)
Packing medium	PACKMED	THS	R	Langual facet K (Møller and Ireland, 2000)
Food contact surface	FDCTSRFC	{THS}	R	Langual facet N (Møller and Ireland, 2000)
Container or wrapping	CONTWRPG	{THS}	R	Langual facet M (Møller and Ireland, 2000)
Storage conditions	STORCOND	{MEM}	O	Storage conditions and duration before arrival at lab
Area of origin	AREAORIG	{THS}	R	Origin of main raw material or area where food was produced if food is a mixed product. Langual facet R
Area of processing	AREAPROC	{THS}	R	Use if different from AREAORIG. Langual facet R (Møller and Ireland 2000)
Area of consumption	AREACONS	{THS}	R	Langual facet R (Møller and Ireland, 2000)
Customary uses of food				
Consumer group label claim	LBLCLAIM	{THS}	R	Langual facet P (Møller and Ireland, 2000)
Specific gravity	SPECGRAV	NUM	O	May also be considered a component. It is the density of the food divided by the density of water at the same temperature. Specific gravity is used to convert to and from standard volumetric or household measures

TABLE A16 (*Continued*)

Property name	Property ID	Data type	Prio	Scope note
Typical serving size	SERVSIZE	NUM	R	In grams
Typical package weight	PACKWGHT	NUM	R	In grams
Typical weight per piece	PIECWHGT	NUM	R	In grams
Frequency and season	FREQSEAS	{STR}	O	How often and in which season is the food preferably consumed?
Place of food in diet	PLACDIET	{STR}	O	How does the food relate to other foods in the diet? Is it a major source of some nutrient?
Cuisine	CUISINE	{STR}	O	Possible future LanguaL facet Q. The special diet a food belongs to (e.g. Mediterranean cuisine) (Pennington *et al.*, 1995)
Sampling and laboratory handling				
Date of sampling	DATSAMPL	DAT	R	When was the sample obtained, purchased, harvested, etc.?
Sampling strategy	SAMPSRAT	{MEM}	R	Brief description of the sampling strategy
Weights of samples	SPLEWGHT	NUM	R	In grams
Place of sampling	PLCECOLL	{STR}	R	Where was the sample obtained, purchased, harvested, etc.?
Number of samples	NRSAMPLE	INT	R	In case of compound sample
Sample handling	SPLEHAND	{STR}	R	General handling of sample before arrival at laboratory, e.g. sample transport
Supplier laboratory of sample	SUPPLAB	FKY	R	Link to organization table (Organization ID)
Date of arrival at laboratory	ARRIVAL	DAT	R	
Laboratory storage	LABSTORE	{STR}	R	Storage conditions in the laboratory before the start of the analytical process
Reason for analysis	REASON	{STR}	R	Context of investigation, e.g. for clinical, comprehensive, control, or contamination study
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A17

Contributing food table

Property name	Property ID	Data type	Prio	Scope note
Food ID	FOODID	FKY	M	Link to the food table, i.e. the aggregate food of the food–food relationship
Contributing food ID	CONFDID	FKY	M	Link to the food table, i.e. the contributing food of the food–food relationship
Amount of Ingredient	AMOUNT	FRC	R	The amount of an ingredient (i.e. a contributing food) may be given as a fraction of the aggregate food
Rank	RANK	INT	R	Often, the amount of ingredients is not known, only their order. In this case, the rank of each ingredient should be given, starting with the most significant ingredient by weight (i.e., 1, 2, 3, …)
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A18

Component table

Property name	Property ID	Data type	Prio	Scope note
Source	SOURCEID	FKY	M	Link to the source table, i.e. the data source where this component is reported
Original component code	ORIGCEPCD	{STR}	R	The component code, ID, or abbreviation used to ident- ify the component in the original publication
Component name	COMPNAME	{STR}	M	The component name in the language given in the attribute Language
Abbreviated component name	ABBREV	{STR32}	O	Maximal 32 characters. Used for applications with limited screen/paper space
Unit	UNIT	THS	M	According to Table 5
Mode of expression	MOEX	THS	M	According to Table 6
Standard classifications				
INFOODS tag name	INFDSTAG	THS	R	See http://www.fao.org/infoods/
EUROFOODS component name	EUFDSNAM	THS	R	Use EUROFOODS list of component names (Schlotke *et al*., 2000)
CAS Registry-Number	CASNR	{STR}	O	As found in the CAS registry file maintained by Chemical Abstract Services

TABLE A19

Contributing component table

Property name	Property ID	Data type	Prio	Scope note
Component ID	COMPID	FKY	M	Link to the component table, i.e. the super-component of the component–component relationship
Contributing component ID	CONCMPID	FKY	M	Link to the component table i.e. the sub-component of the component–component relationship
Weight	WEIGHT	NUM	O	In case of weighted aggregation, a weight or factor conversion can be stored for every sub-component
Profile name	PROFNAM	{STR}	O	

TABLE A20

Method table

Property name	Property ID	Data type	Prio	Scope note
Method headline	METHHDLN	THS	R	According to Schlotke *et al*. (2000)
Method name	METHNAME	{STR}	M	
Scope and general description	GENDESC	{MEM}	R	
Bibliographic reference	BIBREF	FKY	R	Link to the Publication table (Publication ID), i.e. a publication describing the method
Method type	METHTYPE	THS	M	According to Table 8
If an analytical method				
Sample handling	SAMPHAND	{MEM}	R	Includes description of sample preparation, extraction and clean-up at the laboratory

TABLE A20 (*Continued*)

Property name	Property ID	Data type	Prio	Scope note
Analytical details	ANDETAIL	{MEM}	R	Detection procedure, quantification procedure, confirmation procedure, quality control, use of reference material and methods, etc.
Accuracy	ACCURACY	{STR}	R	The closeness of the agreement between the result of a measurement and a true value of the analyte. It may be assessed by the use of reference materials
Applicability	APPLICAB	{MEM}	R	Specify the matrix, concentration range and, for Codex purposes, the preference to be given to "general" methods
Limit of detection	LOD	NUM	R	The detection limit is conventionally defined as field blank + 3σ, where σ is the standard deviation of the field blank value signal
Limit of determination	LODET	NUM	R	As for detection limit except that 6σ or 10σ is required rather than 3σ
Limit of quantification	LOQ	NUM	R	As for detection limit except that typically at least 10σ is required
Precision	PRECISIO	NUM	R	The closeness of the agreement between independent test results obtained under prescribed conditions. The values obtained normally encompass both repeatability intra-laboratory and reproducibility inter-laboratory
Repeatability (intra-laboratory)	REPEAT	NUM	R	The value r below which the absolute difference between the two single test results obtained under repeatability conditions (i.e. same sample, same operator, same apparatus, same laboratory and short interval of time) may be expected to lie within a specific probability (typical 95% and hence $r = 2.8 \times$ S.D. where S.D. is the standard deviation, calculated from results generated under repeatability conditions
Reproducibility	REPRODUC	NUM	R	The value r below which the absolute difference between single test results obtained under reproducibility conditions (i.e. on identical material obtained by operators in different laboratories, using a standardized test method) may be expected to lie within a specific probability (typical 95% and hence $r = 2.8 \times$ S.D. where S.D. is the standard deviation, calculated from results generated under reproducibility conditions
Recovery	RECOVERY	NUM	R	Proportion of the amount of analyte present or added to the test material which is extracted and presented for measurement
Selectivity	SELECTIV	NUM	R	
Sensitivity	SENSITIV	NUM	R	
Specificity	SPECIFIC	NUM	R	The freedom of the analytical procedure from interference effects. It reflects the ability of the instrumentation to measure only the signal of the determined element
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A21

Value table

Attribute name	Attribute ID	Data type	Prio	Scope note
Value ID	VALUEID	KEY	M	These attributes link to food, component and method description, respectively. Such links are mandatory if the value is on the top level of the value hierarchy (see Figure 3) and optional otherwise. This means that values provided by the responsible data source always need to be documented (primary data). Documentation of further data underlying the primary data, however, is optional
Food	FOODID	FKY	M/O	
Component	COMPID	FKY	M/O	
Method	METHID	FKY	M/O	
Best location	BESTLOC	NUM	R	According to Klensin (1992). The value that is considered the best representative according to the decision of the data compiler. Generally, this attribute is mandatory. In some cases, however, it might not be possible to assign a Best Location (e.g. where the distribution shows two clusters of values). In this case, Best Location may be left empty and the reader is referred to the raw data itself. Another possibility is to separate the two (or more) clusters as separate entries in the value table but with the same food and component reference. A third possibility is to consider extra food definitions of the various clusters
Value type	VALTYPE	THS	M	The Value Type is designed to further describe the figure in *Best Location* or to give a qualitative description of the value when no *Best Location* can be given. Use value types as given in Table 7
Quality index	QI	{STR}	R	Result of any systematic quality assessment applied by the data provider. A description of the quality assessment procedure should be given under primary source description
Original source	SOURCEID	FKY	R	Link to source table to document the original source (secondary source) of a value in the case that a third-party value is borrowed or otherwise used within an aggregation. This link is *not* used to document the source represented by the interchange package itself (i.e. the primary source). This is done via food and component description
Date of analysis	DATEANAL	DAT	O	The date when this particular value was analysed
Statistics				
n	N	INT	R	Number of values contributing to the statistic, e.g. analytical replicates, number of samples, number of values from different sources, etc. The other statistical parameters must be based on this number *n*
Mean	MEAN	NUM	R	The mean value of the statistic
Median	MEDIAN	NUM	R	The median value of the statistic
Standard deviation	STDV	NUM	R	Should be used for normal distributions only. Do not mix with standard error Standard deviation $= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Standard error	STERR	NUM	O	Standard error $= \frac{\text{Standard deviation}}{\sqrt{n}}$
Minimum	MIN	NUM	R	The minimal value within the statistic
Maximum	MAX	NUM	R	The maximal value within the statistic
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A22

Percentile table

Attribute name	Attribute ID	Data type	Prio	Scope note
Value ID	VALUEID	FKY	M	Link to the value table, i.e. the value the percentile belongs to
Percentile	PERCENTL	NUM	M	must be >0 and <100
Value	VALUE	NUM	M	The actual value of the percentile
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A23

Statistical value table

Attribute name	Attribute ID	Data type	Prio	Scope note
Value ID	VALUEID	FKY	M	Link to the value table, i.e. the value the percentile belongs to
Value	VALUE	NUM	M	An actual single value
Weight	WEIGHT	FRC	O	In case of weighted aggregations, a weight can be stored for every single value
Remarks	REMARKS	{MEM}	O	Any further remarks

TABLE A24

Contributing value table

Attribute name	Attribute ID	Data type	Prio	Scope note
Value ID	VALUEID	FKY	M	Link to the value table, i.e. the super-value of the value–value relationship
Contributing value ID	CONVALID	FKY	M	Link to the value table, i.e. the sub-value of the value–value relationship
Weight	WEIGHT	FRC	O	In case of weighted aggregations, a weight can be stored for every single sub-value
Remarks	REMARKS	{MEM}	O	Any further remarks

REFERENCES

BGVV (1996). *Konzept, Aufbau und Dokumentation der Datenbank bldsat — Begleitbuch zum Bundeslebensmittelschlüssel.* BgVV-Heft 02/1996, Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin, Berlin.

Burlingame, B. A., Cook, F. M., Duxfield, G. M., and Milligan, G. C. (1995). Food data: numbers, words and images. In *Quality and Accessibility of Food-related Data*, (H. Greenfield, Ed.), Vol. I, pp. 175–182. AOAC International, Arlington, Australia.

Burlingame, B. A., Milligan, G. C., Quigley, R. J., and Spriggs, T. (1996). *FOODfiles Manual.* New Zealand Institute for Crop and Food Research Ltd, Palmerston North.

CIAA (1994). *The CIAA Food Categorisation System, a tool for allocating additives*. CIAA (Confederation of the Food and Drink Industries of the EEC) Document ADD/385/90E rev. 5.

Codex Alimentarius Commission (1989). *Codex Alimentarius*, Abridged Version (B. L. Smith, Ed.), FAO and WHO.

Codex Alimentarius Commission (1993a). *Classification of Foods and Feeds*. Codex Alimentarius, Vol. 2, pp. 150–157. FAO/WHO, Rome.

Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme (1993b). *Pesticide Residues in Food*, Vol. 2. FAO/WHO, Rome.

Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Codex Committee on Food Additives and Contaminants (1995). *Revision of the CIAA (Confederation of the Food and Drink Industries of the EEC) food categories for development of a Codex Food Identification System (CFIS) in regard to the proposed draft Codex General Standard for Food Additives (Prepared by U.S.A.)*. FAO-ESN-CX/FAC 96/10. Rome.

Codex Alimentarius Commission (1996). *Codex Food Categorization System (CFCS) for the General Standard for Food Additives (GSFA)*. Codex Commission on Food Additives and Contaminants. Document no. CL 1996/14-FAC, Part II. FAO/WHO, Rome.

Codex Alimentarius Commission (1997a). *GSC, General Standard for Contaminants and Toxins in Foods*. ALINORM 97/12A Annex V. FAO/WHO, Rome.

Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Codex Committee on Methods of Analysis and Sampling (1997b). *Criteria for Evaluating Acceptable Methods of Analysis for Codex Purposes*. CX/MAS 97/3 and CX/MAS 97/3-Add.1. FAO/WHO, Rome.

Connolly, D. (1997). *XML: Principles, Tools, and Techniques. World Wide Web J.* **2** (4), 1997.

Corkill, M. (1995). *UK Nutrient Databank, Machine-Readable Files*. The Royal Society of Chemistry, Cambridge, UK.

Council Regulation (EEC) No 2658/87 of 23 July 1987 on the tariff and statistical nomenclature and on the Common Customs Tariff. Official Journal of the EEC, No. L256, 07/09/1987.

Favier, J. C., Ireland-Ripert, J., Toque, C., and Feinberg, M. (1995). *Répetoire Général des Aliments, Table de Composition, 2e éd.* Tec & Doc-Lavoisier, Paris.

Greenfield, H. and Southgate, D. A. T. (1992). *Food Composition Data — Production, Management and Use*. Chapman & Hall, London, UK.

Hendricks, T. C. (1992). LanguaL — an automated method for describing, capturing and retrieving data about food. In *International Food Databases and Information Exchange, World Rev. Nutr. Diet* (A. P. Simopoulos and R. R. Butrum, Eds.), Vol. 68, pp. 94–103. Karger, Basel, Switzerland (also see http://food.ethz.ch/langual/).

Herwijnen, E. V. (1994). *Practical SGML*, 2nd ed. Kluwer Academic Publishers Group, Norwell, MA, U.S.A., Dordrecht.

Holden, J. M., Schubert, A., Wolf, W. R., and Beecher, G. R. (1987). A system for evaluating the quality of published nutrient data: selenium, a test case. In *Food Composition Data: A User's Perspective* (W. M. Rand *et al*., Eds.). The United Nations University, Tokyo.

Ireland, J. and Møller, A. (2000). Guidelines for food classification and description in food data banks. *J. Food Comp. Anal.* **13,** 537–546.

ISO 1000 (1992a). SI units and recommendations for the use of their multiples and of certain other units incl. Amendment 1:1998 to ISO 1000:1992.

ISO 2108 (1992b). Information and documentation — international standard book numbering (ISBN).

ISO 3166-1 (1997). Codes for the representation of names of countries and their subdivisions — Part 1: Country codes.

ISO 639 (1988a). Code for the representation of names of languages.

ISO 3166-2 (1998b). Codes for the representation of names of countries and their subdivisions — Part 2: Country subdivision code.

ISO 3297 (1998c). Information and documentation — international standard serial numbering (ISSN).

ISO 8601 (1988d). Data elements and interchange formats — information interchange — representation of dates and times.

ISO/IEC 8859-1 (1998e). Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1.

ISO 9660 (1988f). Information processing — volume and file structure of CD-ROM for information interchange.

ISO 6093 (1985). Information processing — representation of numerical values in character strings for information interchange.

ISO 8879 (1986). Information processing — text and office systems — standard generalized markup language (SGML).

ISO/IEC 646 (1991). Information technology — ISO 7-bit coded character set for information interchange.

Klensin, J. C., Feskanich, D., Lin., V., Truswell, A. S., and Southgate D. A. T. (1989). *Identification of Food Components for INFOODS Data Interchange*. The United Nations University, Tokyo (also see http://www.fao.org/infoods/tags/0tags.htm).

Klensin, J. C. (1992). *INFOODS Food Composition Data Interchange Handbook*. The United Nations University, Tokyo.

Klensin, J. C. (1995). Data identification considerations in international interchange of food composition data. In *Quality and Accessibility of Food-related data*, (H. Greenfield, Ed.), Vol. I, pp. 165–173. AOAC International, Arlington, Australia.

Kohlmeier, L. (1992). The Eurocode 2 food coding system. *Eur. J. Clin. Nutr.* **46** (Suppl. 5), 25–34.

Møller, A. and Ireland, J. (2000). *LanguaL 2000 — thesaurus*. COST Action 99 Report, European Commission, Luxembourg.

Møller, A. and Schlotke, F. (1996). Inventory of European food composition databases. Report to the COST Action 99 Working Group on Data Interchange, http://food.ethz.ch/cost99/db-inventory.htm.

Møller, A. and Saxholt, E. (1996). *Levnedsmiddeltabeller, 4. udgave*. National Food Agency of Denmark, Søborg.

Møller, A. (1992). NORFOODS computer group — food composition data interchange among the Nordic countries: A report. In *International Food Databases and Information Exchange, World Rev. Nutr. Diet* Vol. 68, (A. P. Simopoulos and R. R. Butrum, E.), pp. 94–103. Karger, Basel, Switzerland.

Pennington, J. A. T., Hendricks, T. C., Douglas, J. S., Petersen, B., and Kidwell, J. (1995). International Interface Standard for Food Databases. In *Food Additives and Contaminants*, Vol. 12, pp. 809–820.

Pennington, J. A. T. (1996). Issues of food description. *Food Chem.* **57**(1), 145–148.

Ricketson, S. (1995). International and Australian copyright considerations in data and data compilations. In *Quality and Accessibility of Food-related Data*, Vol. I, pp. 257–273 (H. Greenfield, Ed.), AOAC International, Arlington, Australia.

Schlotke, F., Becker, W., Ireland, J., Møller, A., Ovaskainen, M. L., Monspart, J., and Unwin, I. (2000). *EUROFOODS recommendations for food composition database management and data interchange*. COST Action 99 Report, European Commission, Luxembourg.

Schlotke, F. (1996). Using Internet services to improve international food data exchange. *Food Chem.* **57,** 137–143.

Shoshani, A. (1991). Properties of statistical and scientific databases. In *Statistical and Scientific Databases* (Z. Michalewicz, Ed.). Ellis Horwood, New York.

Stewart, K. K. (1988). Editorial: are they different. *J. Food Comp. Anal.* 1, 2, 103.

Trichopoulou, A. and Lagiou, P. (1997). *Methodology for the exploitation of HBS and results on food availability in five European countries.* European Commission, EUR 17909 EN, Luxembourg.

Truswell, A. S., Bateson, D., Madafiglio, D., Pennington, J. A. T., Rand, W. M., and Klensin, J. C. (1991). INFOODS guidelines for describing foods to facilitate international exchange of food composition data. *J. Food Comp. Anal.* **4,** 13–38.

Unwin, I. and Becker, W. (1996). The component aspect identifier for compositional values. *Food Chem.* **57,** 149–154.

Unwin, I. (1999) *Food Table Input*: http://www.ianunwin.demon.co.uk/foodtab/.

USDA-ARS Nutrient Data Laboratory (1998). *USDA Nutrient Database for Standard References, Release 12*. Documentation. Beltsville Human Nutrition Research Center, Riverdale, MD (see http://www.nal.usda.gov/fnic/foodcomp/).

World Customs Organization (1996). *Harmonized commodity description and coding systems, Explanatory notes. 2nd edn.* World Customs Organization, Brussels.

World Wide Web Consortium (1999). *XML*, http://www.w3.org/XML/.