ELSEVIER

PII: S0308-8146(96)00115-X

Food Chemistry, Vol. 57, No. 1, pp. 137–143, 1996

0308-8146/96 $15.00+0.00

Using Internet services to improve international food data exchange

Florian Schlotke

Institute of Scientific Computing, ETH Zurich, 8092 Zurich, Switzerland

Based on the World Wide Web (WWW) means to provide on-line interactive access, both nationally and internationally, to a Swiss Food Composition Database (SWIFD) are investigate. The architecture of an information-server and a self-explanatory interface are described that allows the user to conveniently view and copy data as well as information on data quality and food description. Using the concept of federated databases a network of distributed information-servers connected by the Internet is outlined. It was found that the technical aspects are minor compared with the semantic and legal (copyright) problems that arise with international data exchange. As a consequence, additional collaborative efforts are required, particularly in the field of standardized exchange formats and food descriptors. Copyright © 1996 Elsevier Science Ltd

INTRODUCTION

Because information on the composition of foods is important in a variety of applications (Rand *et al.*, 1987), many countries have established organizations that analyse foods and compile food composition databases. There are also international organizations interested in data from different national databases (e.g. the International Agency for Research on Cancer—IARC). Limited resources have led to the common practice of substituting one's own analysis with food composition data from foreign sources (Rand *et al.*, 1991; Greenfield & Southgate, 1992) increasing the importance of *international data exchange*.

Four aspects must be considered when exchanging data, namely *organizational*, *logical*, *physical* and *legal aspects*.

1. Organizational aspects: which organizations are involved in data interchange and how do they interact? Who is responsible for what type of data? Who sets and maintains standards?
2. Logical aspects: how are data and its quality described? How are data structured and formatted?
3. Physical aspects: what medium is used to transfer the data?
4. Legal and financial aspects: is there a copyright on the data? How are users of data charged? Who is responsible for the correct use of data?

Much effort has been put into the development of organizational networks such as the International Network of Food Data Systems (INFOODS) and its regional data centres (Klensin, 1992) as well as models to integrate food composition databases with other databases covering food consumption or food labelling (Soergel *et al.*, 1992). Several efforts are under way to cover the logical aspect (Klensin *et al.*, 1989; Truswell *et al.*, 1991; Hendricks, 1992; Kohlmeier, 1992; Unwin & Becker, 1995). No international consensus or standard has been established, however, and food composition data are not yet regularly exchanged. The physical aspect of data exchange is the topic of this article: we propose to publish food composition data and all related descriptive data on the Internet. We describe how Internet communication services are being used for a prototype system in Switzerland.

The Internet is already widely used in other scientific fields to exchange data such as laboratory results, research programme information, computer software and even electronically published scientific journals. The Internet has also been used to support collaborative scientific work (Fox & Lancaster, 1994). The US Department of Agriculture (USDA) started to provide food composition data on the Internet (the current address is http://www.nal.usda.gov/fnic/foodcomp).

Legal and financial aspects have previously been discussed in detail by Ricketson (1995). We will briefly cover these issues and their relationship to electronic data interchange in the last section.

LAY USERS VS SCIENTIFIC USERS

We distinguish between two types of users of a food composition database published electronically by a national data compiler or an international organization: *lay users* and *scientific users*. They differ with respect to the type of data they need and the way they use the data.

The lay user needs a well compiled, complete database for use in a specific application (e.g. some dietary software). He or she trusts the data compiler who designed the database for the user's purposes in respect to the adequacy, correctness and accuracy of the data.

The scientific user is interested both in the actual food composition data and in the data describing the food items, the components and the analytical values. Without this additional descriptive data, subsequently called *meta-data*, it is not possible to interpret the values and their quality correctly. A scientific user may be a data compiler or a scientist comparing his or her own analysis with other sources.

The scientific user is probably most interested in online *browsing* and *search facilities* within larger and more detailed databases, whereas a lay user prefers to *copy* complete food composition tables, ideally accompanied by some application software, for further processing. Both actions need to be supported in a convenient and comprehensive way by the data distributor. We consider a browsing facility for scientific users the highest priority.

COMMUNICATION: CLIENTS AND SERVERS

During the last decade, the metaphor client–server became increasingly popular to illustrate the relationship between partners of a communication. Depending on different contexts the terms client and server may be used to designate people: a national data compiler as the server and a dietician as the client; computers: a server-computer holding a database and a client-computer accessing this database via a network; programs (or processes) running on computers: a WWW-server process and a WWW-client program such as Mosaic.

There is a slight but important semantic shift when we turn from the general terms of sender and receiver to the model of clients and servers: the active role turns from the sender to the client. The client looks for what he/she needs, instead of passively receiving whatever the sender sends. In other words, world-wide access to food composition data will become more efficient and interactive if some automated server system on a public computer network (e.g. the Internet) is provided.

We call such an automated server system an *information-server* and define it as a computer system (i.e. a collection of hard- and software components) that enables clients, in any of the three senses mentioned above, to access information and data via some computer network. The information and data may be of textual or any audio-visual form; computer software is covered by this.

INTERNET SERVICES

The Internet is a world-wide computer network of independent sub-networks and single computers that use a standardized communication protocol (i.e. TCP/IP) for data transmission. Since it first started, more than two decades ago, a collection of general purpose services have been developed culminating in the World Wide Web (WWW) (Berners-Lee *et al.*, 1994). Most of these services can be used to build an information-server for food composition data. We use the WWW for our project because it combines and extends other services that emerged earlier.

Electronic mail (e-mail)

Electronic mail is mostly used to exchange plain text messages (i.e. ASCII-format) between people on the Internet or other computer networks (e.g. CompuServe, Bitnet). Messages are buffered in a mailbox on the target machine until read by the receiver. Instead of reading e-mail personally it is also possible to automatically parse and process the message using a program, thus providing a system that could answer queries sent by e-mail from remote users. The answer is sent back again using e-mail. This would not be a perfectly interactive system. Moreover, e-mail is not a powerful tool for transmitting other data than plain text (e.g. graphics).

File transfer protocol (FTP)

FTP is a service to copy any type of file (text, graphics, programs, etc.) from one machine on the Internet to another. Normally this requires the user to have an account (i.e. explicit access privileges) on both the source and the target machine. Public FTP-servers can be anonymously contacted to copy any file that is offered by the server machine. With anonymous FTP, however, the user must know exactly what files he/she needs and where these files are stored. Given a particular file, the user has no influence on the type or amount of data within this file. There is a lack of interaction in the concept of FTP.

Telnet

The Telnet service is used to set up interactive information systems. A user can log-on to a remote system using a text terminal (or terminal emulation). The server system has full control of the user interface and the configuration of a session. Public library systems often use the Telnet service. Because of the limited screen size of text terminals, this concept is rather inconvenient for browsing large amounts of data. Moreover, the user has to cope with a different user interface (or set of commands) for each Telnet-server.

World Wide Web (WWW)

The WWW is a globally distributed hypertext system

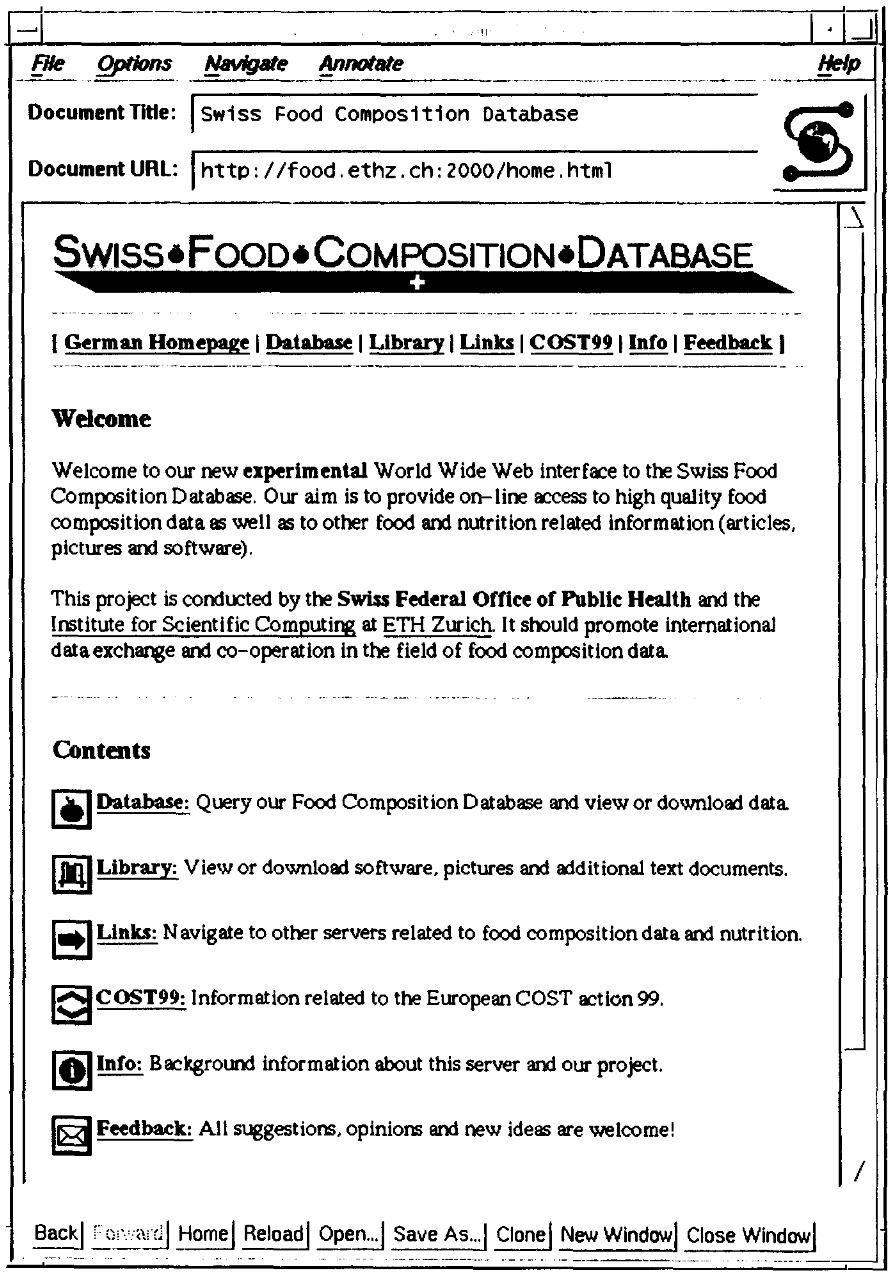

Fig. 1. Mosaic screen shot showing the 'SWIFD' information-server's entry page.

introduced by CERN (European Centre for Nuclear Research). A hypertext is a document containing specially marked words, or keywords, that refer to other documents. In case of the WWW these documents may be located on any other machine on the Internet. A WWW document may also include references (or links) to e-mail services, FTP- and Telnet-servers. The possibilities to format a document using the *Hyper Text Markup Language* (HTML) and to include graphics made the WWW a popular service on the Internet (Fig. 1). Many commercial and government organizations and universities are publishing information on the WWW (Krol, 1994). A WWW-client program (sometimes called WWW-browser or Web browser) is needed to display HTML-documents distributed by a WWW-server. There are different WWW-client programs available for all major computer systems. Our discussion refers to Mosaic, a WWW-client developed at the National Centre for Supercomputing Applications at the University of Illinois (NCSA) (Dougherty *et al.*, 1994).

Another reason to implement our food composition data information-server using the WWW is the possibility to create forms with HTML. Forms allow the user to input text, to specify options and to start separate programs on the WWW-server machine using the *common gateway interface* (CGI) feature within the WWW concept as described below for the SWIFD information-server architecture. Such a program computes a result, depending on the user's input. The WWW-server relays the result back to the user's WWW-client. Thus, we have a means to anonymously query a remote database server in an interactive way.

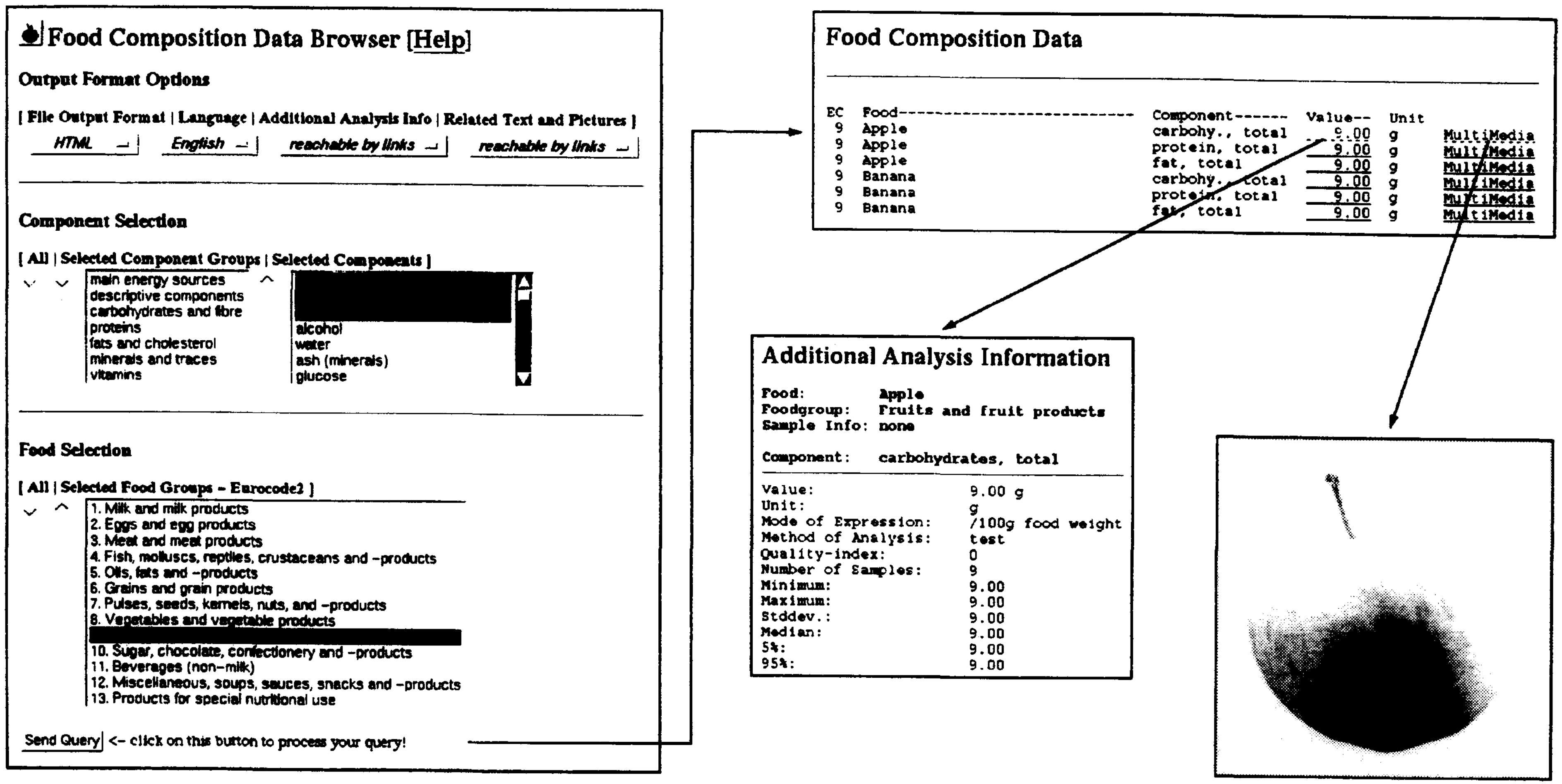

Fig. 2. The food composition database query interface, the result of a sample query and dyanamically linked documents.

THE 'SWIFD' INFORMATION-SERVER

The experimental prototype of an information server for a future Swiss Food Composition Database (SWIFD) built and operated at our Institute (http://food.ethz.ch:2000/) has been designed to meet the following requirements: the server should allow scientific as well as lay users to get their specific data. The server should also be used to distribute information related to food and nutrition such as articles, software tools or a nutrition help desk. References to other related servers on the WWW guide the user to further information (Fig. 1).

The lay user

We meet the needs of a lay user by providing access to predefined files via the WWW and FTP. We plan to offer ready-to-use food composition tables together with additional notes on food and component description and advises for proper usage. The tables will be formatted for import and further processing in standard spreadsheet or database systems. The user will also be able to copy data together with software tools to browse, print and process the food composition tables on personal computers.

The scientific user

The scientific user has interactive access to food composition data and its related meta-data. With the help of the food composition data browser shown in Fig. 2 a query can be specified. The main sections of this form are the food selection, the component selection and a series of output format options. The result of the query can be printed using the INFOODS interchange format (Klensin, 1992) and the INFOODS tag-name system for component description (Klensin *et al.*, 1989). When the HTML format is selected, additional information can be dynamically linked to the result file. These links provide a path to further food, component and analytical value description as well as to related text documents and pictures (Fig. 2). The referenced files are reached by a mouse-click on the underlined words or figures. A tabular presentation of additional analysis information is also available. The exact content of this table of meta-data, however, is still the subject of further discussions. The language option controls the presentation of food, component and value descriptions. This feature is particularly helpful in Switzerland where a server must cope with four official languages.

Within a query it is possible to specify all available components, to select specific component groups (e.g. vitamins) or to select an arbitrary set of components. Foods can be selected by food groups according to the Eurocode2 system (Kohlmeier, 1992). Selecting foods based on a set of Langual codes (Hendricks, 1992) or a search string or keyword to match with a food's name is planned for future implementation.

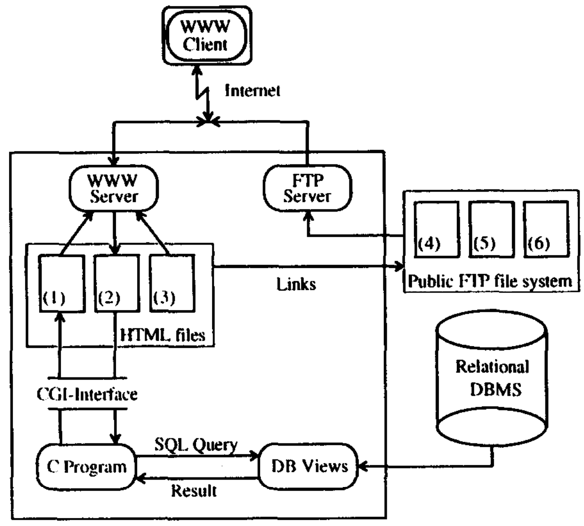

Fig. 3. Information-server architecture.

The server's architecture

Figure 3 gives a schematic overview of the information-server's architecture. The user interface consists of a set of files written in HTML. These files are interpreted and displayed by the WWW-client. Some of the HTML-files are read-only [(3) in Fig. 3], some contain forms for data input (2) and some are dynamically created (1) by programs invoked by forms via the common gateway interface (CGI). Each HTML-file may include references to additional text-files (4), graphics-files (5) or software archive-files (6) distributed by the FTP-server. The food composition data datas stored in a relational database management system (DBMS) connected to the information-server. The programs collect the data via SQL queries (Structured Query Language). The database is accessed indirectly through predefined views. The database itself is read-only: the user (i.e. the WWW-client) is in no position to change the data. This can only be done by the data compiler with a separate program not discussed in this paper. The database only views data that have been approved for publication.

Technical notes

The following paragraph briefly reviews the tools used to implement the server's components. Because WWW technology is platform-independent other configurations are possible. We use a UNIX workstation, which is connected to the Internet via a campus-wide local area network. The food composition data are stored in an Oracle database which is accessed using C programs with embedded SQL statements. NCSA's WWW-server v1.3 is used for data distribution. This server is available as public domain software. A brief overview of HTML and the WWW concept, as well as information on getting WWW-related software, is given in Vetter *et al.* (1994) and Dougherty *et al.* (1994). A good starting point for information on setting up a WWW-server as well as other general technical specifications is the homepage of the World Wide Web consortium at http://WWW.w3.org.

FUTURE WORK

A network of federated databases

Since the idea of a centralized food composition database at the global or European level has been dropped, INFOODS developed the concept of a network of autonomous regional data centres. Each data centre might use a different database management system or database schema design. Data interchange, however, is done in a standardized manner (Klensin, 1992). It is proposed that this concept by is implemented by connecting distributed information-servers on the Internet to a federated database system (FDBS).

A federated database system is a collection of co-operating distributed database systems that are autonomous and possibly heterogeneous. The level of *distribution* (number of places where data is stored), *heterogeneity* (differences in the type of DBMS used or differences in the semantics of data) and *autonomy* (freedom of the participants to choose their proper database design and to decide what data and functionality to share) characterizes an FDBS (Seth & Larson, 1990).

In the context of INFOODS we are faced with a high level of distribution, heterogeneity and autonomy. With the Internet we can hide the distribution of data. Heterogeneous DBMSs are hidden by the WWW interface. Heterogeneous data models and data schemas can be integrated if the problem is reduced and a small set of fixed queries are defined that cover most of the user's needs. Each information-server maps the queries to the underlying DBMS. Autonomy is guaranteed because each data centre is in charge of its own information-server, may decide which data to export (DB Views in Fig. 3) and has full control over its database: data is read-only for external users.

There are two possibilities to interact with the FDBS depending on how tight the different databases are coupled (Seth & Larson, 1990). In a loosely-coupled FDBS a user queries each server separately. Different user interfaces or query languages and the need to formulate multiple queries make this approach unattractive. The second possibility is to build a software component that is able to distribute a query to the available servers, to merge the results, tagging them with source information, and to present the results in a uniform way. Such a component can be called a mediator (Wiederhold, 1992). The mediator itself is accessible on the Internet through a WWW-server. It can also be part of one of the participating information-servers (Fig. 4). We now have a tightly coupled FDBS with the mediator's set of possible queries representing the global conceptual schema of the FDBS. It will be manageable to build appropriate mediators and to couple different servers if data presentation is standardized.

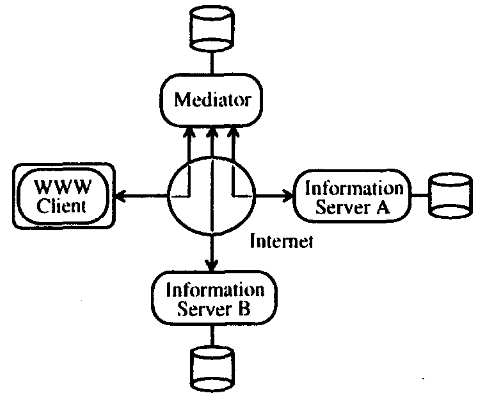

Fig. 4. Coarse architecture of a tightly coupled federation of food composition information-servers.

The FDBS is presented to the user by adding an option field to the food composition data browser proposed in Fig. 2. The user will get an answer from all the servers he or she selected out of a given set of possible information-servers within the network.

DISCUSSION AND CONCLUSIONS

The proposed methods and their restrictions affect the four aspects introduced at the beginning as follows.

Organizational aspects

A network of distributed information-servers is an appropriate representation of the network of regional data centres proposed by INFOODS. It depends on the design of mediators, how data requests are handled and delegated to corresponding servers. Because of reduced publishing time, the Internet can also be used as an efficient public platform to develop and enforce international standards required for data interchange. INFOODS provides an example: an up-to-date list of the official tag-names is published on a WWW-server (http://WWW.crop.cri.nz/crop/infoods/infoods.html).

Logical aspects

A major requirement is the possibility to distribute meta-data. The proposed information-server architecture allows the presentation of meta-data in different ways: in tabular form, as a tagged file format (INFOODS exchange format) or implicitly by linking the meta-data to its corresponding food composition data using hypertext. In order to fully exploit the possibilities of data interchange, further collaborative research at the international level is needed to develop a tagged file format that is fully compatible to international standards (i.e. ISO 8879/SGML) and that covers all aspects of food, component and analytical value description. In addition, standardization of international codes and vocabulary (e.g. LanguaL) needs to be enforced. More software tools are needed for these systems to be used efficiently.

Physical aspects

Compared to traditional mail and fax facilities, the Internet and the WWW are a fast, flexible and user-friendly means of communication. They proved to be a suitable electronic medium to efficiently publish food composition data on an international basis. Most of the software and tools to implement an information-server are available in the public domain. Not all addressed users, however, have access to the Internet. Because of the enormous growing rate of the Internet and related technologies, it is reasonable to assume that this is only a temporary problem.

Legal and financial aspects

Regardless of the medium used for data transmission, legal aspects may pose more difficulties for data interchange than the other aspects. In the context of electronic publication, two strategies to protect the copyright of data providers are typically discussed. The first is to attach a copyright note to every set of data transmitted to the user and to trust the user's honesty to use the data according to the stated conditions. A more radical strategy is to only permit on-line database access to registered users who have paid some copyright fee in advance. The user then gets an identification and a password. A variation of this approach is to offer the meta-data as public domain and to only restrict access to the actual analytical values. This allows the user to look at the type of data in advance.

The Internet and related technologies are still under development. The emerging commercial use of the Internet will lead to more sophisticated techniques that address legal and financial aspects in the future (e.g. electronic money).

Because food composition data that are published on the Internet are easily available, some people might uncritically use it in applications not intended by the data compiler. Therefore complete, high-quality meta-data should be provided and the same rigid quality assurance and editing procedures used with printed publications must be applied in electronic publishing.

ACKNOWLEDGEMENTS

The author would like to thank Ove Hiltwein and Dieter Richter who did most of the programming of our information-server; and Hans Hinterberger, Ian Unwin and the anonymous referees for their valuable comments. This work has been supported in part by the Swiss Federal Office of Public Health under contract No. 316.95.0516.

REFERENCES

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F. & Secret, A. (1994). The World-Wide Web. *Commun. ACM.* **37,** August, pp. 76–82.

Dougherty, D., Koman, R. & Ferguson, P. (1994). *The Mosaic Handbook for the X Window System.* O'Reilly and Associates, Sebastopol, California.

Fox, P. T. & Lancaster, J. L. (1994). Neuroscience on the Net. *Science,* **266,** 994–996.

Greenfield, H. & 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,* Vol. 68, eds A. P. Simopoulos & R. R. Butrum. Karger, Basel, Switzerland, pp. 94–103.

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

Klensin, J. C., Feskanich, D., Lin, V., Truswell, A. S. & Southgate D. A. T. (1989). *Identification of Food Components for INFOODS Data Interchange.* The United Nations University, Tokyo, Japan.

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

Krol, E. (1994). *The Internet User's Guide and Catalog,* 2nd edn. O'Reilly and Associates, Sebastopol, California.

Rand, W. M., Pennington, J. A. T., Murphy, S. P. & Klensin, J. C. (1991). *Compiling Data for Food Composition Data Bases.* The United Nations University, Tokyo, Japan.

Rand, W. M., Windham, C. T., Wyse, B. W. & Young, V. R. (eds) (1987). *Food Composition Data: A User's Perpective.* The United Nations University, Tokyo, Japan.

Ricketson, S. (1995). International and Australian copyright considerations in data and data compilations. In *Proceedings of the 1st International Food Data Base Conference 1993,* ed. H. Greenfield. AOAC International, Arlington, USA, pp. 257–273.

Seth, A. P. & Larson, J. A. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous database systems. *ACM Comput. Surv.,* **223,** 183–236.

Soergel, D., Pennington, J. A. T., McCann, A., Holden, J. M., Smith, E. & Wiley, C. (1992). A network model for improving access to food and nutrition data. *J. Am. Diet. Ass.,* **92,** 78–82.

Truswell, A. S., Bateson, D., Madafiglio, D., Pennington, J. A. T., Rand, W. M. & Klensin, J. C. (1991). INFOODS guidelines for describing foods to facilitate international exchange of food composition data. *J. Food. Comp. and Anal.,* **41,** 18–38.

Unwin, I. & Becker, W. (1995). Component aspect identifier—a tool for handling food component information in a food database management system. Report No. 9/95. National Food Administration, Uppsala, Sweden.

Vetter, R. J., Spell, C. & Ward, C. (1994). Mosaic and the World-Wide Web. *IEEE Comput.,* **27** October, 49–57.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Comput.,* **25** March, 38–49.