Journal of Medical Internet Research

Contents

The SGML Standardization Framework and the Introduction of XML

Walter Fierz[1], MD; Rolf Grütter[2], DVM

[1]Institute for Clinical Microbiology and Immunology, University of St. Gallen, Switzerland

[2]Institute for Media and Communications Management, University of St. Gallen, Switzerland

Corresponding Author:
Walter Fierz, MD
Institute for Clinical Microbiology and Immunology
Frohbergstrasse 3
CH-9001 St. Gallen
Switzerland
Phone: +41 71 494 37 40
Fax: +41 71 494 37 85
Email: walter.fierz@gd-ikmi.sg.ch

Abstract

Extensible Markup Language (XML) is on its way to becoming a global standard for the representation, exchange, and presentation of information on the World Wide Web (WWW). More than that, XML is creating a standardization framework, in terms of an open network of meta-standards and mediators that allows for the definition of further conventions and agreements in specific business domains. Such an approach is particularly needed in the healthcare domain; XML promises to especially suit the particularities of patient records and their lifelong storage, retrieval, and exchange. At a time when change rather than steadiness is becoming the faithful feature of our society, standardization frameworks which support a diversified growth of specifications that are appropriate to the actual needs of the users are becoming more and more important; and efforts should be made to encourage this new attempt at standardization to grow in a fruitful direction. Thus, the introduction of XML reflects a standardization process which is neither exclusively based on an acknowledged standardization authority, nor a pure market standard. Instead, a consortium of companies, academic institutions, and public bodies has agreed on a common recommendation based on an existing standardization framework. The consortium's process of agreeing to a standardization framework will doubtlessly be successful in the case of XML, and it is suggested that it should be considered as a generic model for standardization processes in the future.

Introduction

Extensible Markup Language (XML) Version 1.0 was endorsed as a Recommendation by the World Wide Web Consortium (W3C) in February 1998 (http://www.w3.org/TR/1998/REC-xml-19980210). W3C is a consortium assembling representatives from many of the leading IT companies, academic institutions, and public bodies [1]. Following W3C's recommendation, all major software companies refocused their development strategies to include XML, and many are now offering products implementing the recommendation. These products range from XML parsers implemented in different programming languages, to sophisticated XML-oriented database applications (for a list of tools that support XML see [2]). XML is also supported by the recent version of Microsoft Internet Explorer and will be supported by the next version of the Netscape browser. The speed of general acceptance and widespread use recalls the introduction of the programming language Java a few years ago and - although public awareness has partly shifted to other issues - imagine where we would be today without it!

The objective of this report is to provide an overview of XML and associated standards. Particular emphasis will be given to its applications in the healthcare sector. This overview will include the history of XML and elucidate the factors enabling the rapid diffusion of XML. In addition, emphasis will be put on the standardization process at the two levels involved: the first includes the acceptance of XML as a syntactical specification; the second considers the agreement upon semantic

XSL•FO
RenderX

conventions by specific groups of users, whereby the Document Type Definitions (DTDs) provide the basis for those agreements.

## XML Standardization Framework

A Standardization Framework

Information technology (IT) standards usually put some constraints on ways that human or computer agents interact, in order to allow collaboration on a common basis. In this report, we would like to concentrate on an additional function of a certain type of standards that might be called meta-standards. Meta-standards specify how other standards can be defined. In principle, one can imagine a whole hierarchy of meta-standards and standards that have their application at different levels of specificity. Corresponding to this kind of standards tree, there are different levels of standard bodies responsible for the specifications of the standards, with the scale reaching from the International Standards Organization (ISO) at the top, via various levels of academic and industry initiatives, ultimately down to the level of two communicating individuals who agree on some mutual way of expressing and handling information objects. An additional type of standard is created to serve the function of specifying ways to translate between different standards, in order to allow the transformation of information objects obeying one standard into objects following a different standard. This type of standard can be called a mediator. Again, there are meta-standards that specify how to define specific mediators.

A system of meta-standards and mediators is what we propose to call a standardization framework. Such a system is able to function in diversified areas of IT, and has the potential of being flexible and adaptable to change. Furthermore, the openness of such a system enables rapid development in growing fields. At a time when change rather than steadiness is becoming the

faithful feature of our society, standardization frameworks which support a diversified growth of specifications that are appropriate to the actual needs of the users are becoming more and more important; and efforts should be made to encourage this new attempt at standardization to grow in a fruitful direction. With time, rules might develop that define how such an open system works best. Here, we would like to give an example of a success story of such a standardization framework in the IT area.

SGML/XML: The Growth of a Standards Tree

The last ten years have seen the tremendous development of a hierarchy of standards and quasi-standards that is based on a common root, the Standard Generalized Markup Language (SGML). Particularly, one application of SGML, Hypertext Markup Language (HTML), has revolutionized the development of the World Wide Web (WWW), the way that electronic documents are interchanged on a global scale. A second revolution is on the horizon: another child of SGML, XML, will profoundly transform the way electronic data will be exchanged. The secret behind this second revolution can be summed up by XML's eponymous attribute, "extensible." While extensibility was already a property of SGML, the improved capabilities of XML have enabled the growth of a family of sub-standards that may be flexibly adapted to the needs of particular domains of users. Additionally, it is conceivable to create ad hoc conventions for data exchange between single individuals who do not belong to a pre-defined group. In this view, SGML/XML has not only revolutionized document and data exchange; perhaps even more importantly, it has moved the vision of standards away from a fixed, centralized, and authoritarian paradigm to a libera,l adaptable system of quasi-standards that simply fulfills the actual needs of communication while adhering to a common framework.

XSL•FO

RenderX

Figure 1. The history of SGML/XML and related standards
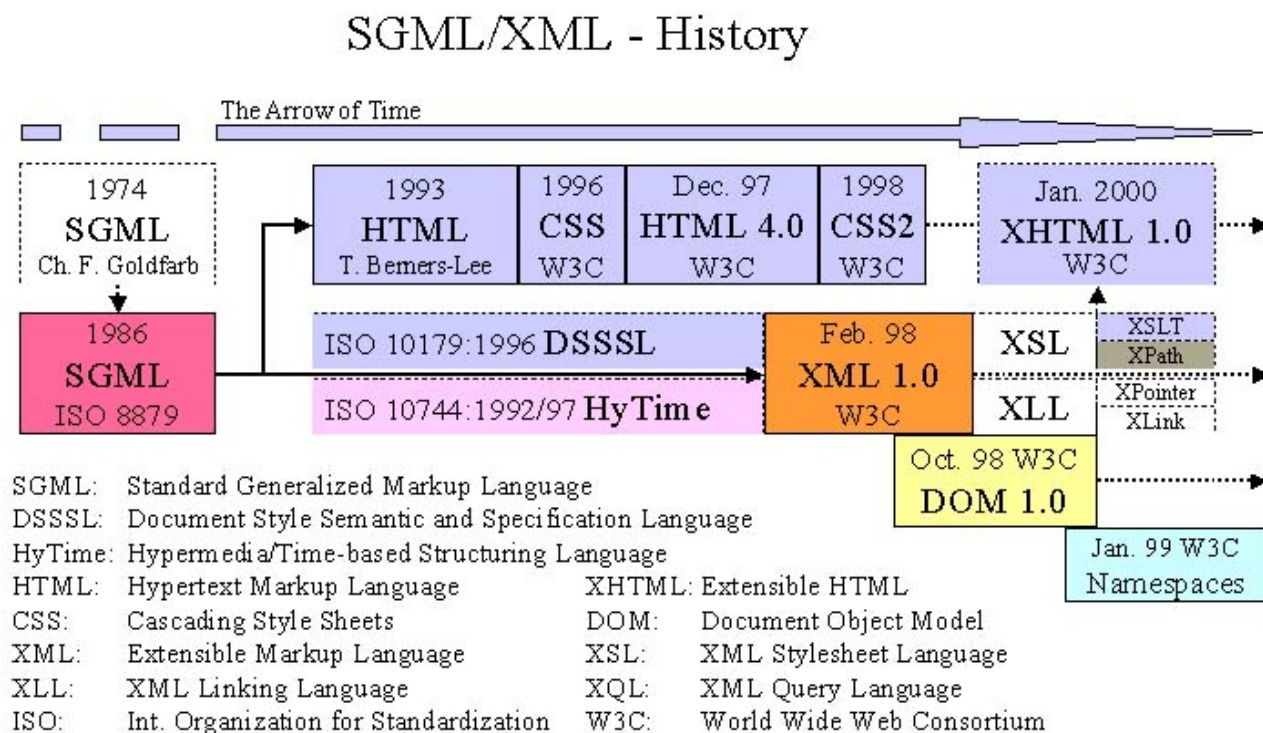
To quote Liora Alschuler:

The central problem of information exchange is the tension between: local specialization and global generalization. In other words, "Why can't everyone just agree to do things my way?" is not the right question. The right question is: "How can we impose minimal constraints on local practice yet...ensure that senders and receivers can share meaning where meaning is, indeed, shared? [3]"

In the following sections, the history of the growth and unfolding of the SGML/XML tree of standards will be sketched, and some of its branches will be explained in more detail. Particular reference will be made to its use in the healthcare domain. Figure 1 summarizes the history of SGML/XML and related standards and recommendations.

SGML (ISO 8879:1986)

The general idea behind SGML is to provide rules that are able to define what the content of a document is, rather then what the document looks like. The separation of the content and structure of a text from the style information used for its rendering in an output device is an important precondition for the document not only to be accessible with heterogeneous software systems, but also to survive in an environment of constantly changing software tools. Inadvertently, this concentration of SGML on the content and structure of a text has made it possible to use SGML not only for documents, but also for data in general, a development that is just beginning now. Furthermore, the origin of SGML in the publishing area has led to a syntax that is not only machine-interpretable but at

least in principle also human-interpretable. This, again, is not unimportant when considering the longevity of certain documents.

The technique that is used by SGML is Markup. Simply speaking, a piece of text or data is enclosed between two tags that "mark it up" as something defined. Such markup allows machine agents to to manipulate the content and to assign a semantic meaning to the piece of data. SGML provides the syntax to define these tags and the rules that govern the use of the markup tags. These definitions are placed either in the prolog of the document or in a separate document called a Document Type Definition (DTD). The overall structure defined by the rules of SGML is that of a tree of elements with one single root element. Each element-type is defined in the DTD by its tag-name and its attributes. The rules given in the DTD govern the possible occurrences of the elements in the tree, and the possible attribute values. Apart from elements and their attributes, the DTD might also define entities. These are character strings that stand for something else, such as longer strings, special characters, groups of elements, or items stored in an external file like graphics or text fragments.

Historically, SGML emerged in the late 60's in the publishing industry, which had a need to establish generic typesetting codes that would allow the text of a document to be manipulated in different text processing systems. One of these approaches was the Generalized Markup Language developed by Goldfarb, Mosher, and Lorie at IBM. A key concept was the DTD, which allowed the construction of markup rules for specific applications and the use of parsers to validate the syntax of the document markup for its appropriateness to the particular

application. Another impetus came from a committee of the Graphic Communications Association (GCA) that created GenCode to standardize typesetting codes. With a collaborative international effort in the 70's under the technical leadership of Charles F. Goldfarb, SGML was developed into an international standard which was adopted by the International Organization of Standards as ISO 8879 in 1986. It has been proven since to be a robust, stable standard that has led to the introduction of manifold applications in the form of further standards, quasi-standards, and academic as well as industry initiatives.

Since only a very short overview over the basic principles of SGML can be given here, we refer for further details to some books dedicated to describing SGML and its applications [4-6].

SGML Applications

The openness and extensibility of SGML stems from the basic concept that the DTD is constructed by the user, or group of users, of the standard. In this sense, SGML is a syntax that allows the definition of further standards. Each DTD defines the particular markup structure that is allowed in a document, and builds a basis by which semantic meaning is conferred on the marked-up content.

Industry-Standard DTDs

Some of the major applications of SGML, in form of DTDs, have been created by industry groups to facilitate document processing and exchange. Table 1 gives a short overview (compiled from [7]).

Table 1. Industry-Standard DTDs

Name	Description	URL / Public Identifier
ISO 12083:1993	International standard defining three DTDs: - books - serial publications - individual articles	ISO 12083:1993//DTD Book//EN ISO 12083:1993//DTD Serial//EN ISO 12083:1993//DTD Article//EN
DocBook	Consortium standard (OASIS) for publishing books with additional features for computer and software documentation	-//OASIS//DTD DocBook V3.1//EN [8]
Text-Encoding Initiative (TEI)	Book-oriented DTD for research-oriented applications. Subset for simpler applications: TEI-Lite	-//TEI//DTD TEI Lite 1.0//EN [9]
MIL-STD-38784 CALS	Book-oriented DTD for technical documents. Known for its table model (CALS) that became a de facto standard for tables.	-//USA-DOD//DTD MIL-STD-38784 AMEND1//EN [10]

Despite its sound and open design, SGML was for some time only used in publishing, mainly by large-scale government and industry enterprises that had to deal with complex documents. It was only when a relatively simple DTD of SGML was combined with a linking mechanism and coded into an Internet-based hypertext application that SGML got into widespread use in form of Hypertext Markup Language (HTML).

HTML and the World Wide Web

As the global public gained access to the former ARPAnet, the Internet evolved; several information services evolved along with it, with the multimedia, interactive WWW gaining the favor of users. In accordance with the hypertext paradigm [11], which refers to a non-sequential linking of information objects, the idea was (and still is) to author and exchange hyper-linked documents. For the implementation of this idea, the SGML standardization framework proved to be an excellent foundation.

The birth of HTML took place 1989 at CERN, where it was developed by Tim Berners-Lee to share hyperlinked text documents within the CERN European Nuclear Research Facility. The new SGML application was made available openly on the Internet, and it soon became very popular and rapidly developed into the WWW as it is well-known and used all over the world today. To be precise, HTML was, at its beginning, not a strictly-conforming SGML application; its development was mainly driven by numerous programmers on the Internet

and by the competing browser vendors. In this sense, HTML developed rather like a de facto standard. Fortunately, the World Wide Web Consortium (W3C), the Web's standards body, managed to get HTML under control and produced DTDs for formal description of each new version of HTML. The first SGML-compliant version was HTML 2.0.

Another departure of HTML from the principals of SGML was perhaps more serious. HTML concentrated mainly on the layout of a document rather then on its content. This development went so far that element tags and attributes were used to directly define formatting instructions. Thus the basic SGML concept of separating content and structure from format was violated. Efforts were made in 1996 by the W3C to halt this development with the introduction of Cascading Style Sheets (CSS). Style sheets have from the beginning been used with SGML to define the formatting of structured text (see section DSSSL). The new versions of HTML browsers are now able to deal with CSS.

HTML is now in its fourth version, which was released by the W3C in December 1997 [12] . The second version of CSS has passed as a W3C Recommendation in May 1998 [13]. Further development of HTML will likely go into the direction of XML (see below), i.e. it has been reformulated as an application of XML and is called XHTML (see below).

HyTime (ISO/IEC 10744:1997)

Hypermedia/Time-based Structuring Language (HyTime) [14] is an SGML application that specifies a way in which logically

XSL•FO
RenderX

connected or time-related information can be described within the framework of SGML. One of its origins is rooted in the attempt to create a language to describe music. However, it can be used for any type of linking between related or time-based information objects.

The main goal of HyTime is to provide rules for linking information objects and scheduling within finite coordinate spaces. In different HyTime modules, methods for addressing locations and methods for hyperlinking between those addresses are defined. Other HyTime modules contain facilities for scheduling and rendition of events within event schedules. Information about HyTime can be found at the site of the HyTime Users' Group [15].

The design principle of HyTime is a meta-DTD, a kind of template called *architectural form* or *enabling architecture*. HyTime lets the user of the standard define his or her own DTDs and relate them to the architectural form with a specific *Hytime* attribute. HyTime-aware applications can recognize these attributes and apply to the corresponding elements any specific processing designed for the HyTime template. Annex A.3 to the second edition of the HyTime standard (ISO/IEC 10744:1997) contains the definition of architectural forms: "Architectural Form Definition Requirements (AFRD)" [16], whereas Annex C standardizes the meta-DTD formalism used for architectures. A general description of SGML architectures by Steven R. Newcomb can be found at [17].

Architectural Forms

HyTime itself is the first, pioneering SGML architecture, but other architectures can be constructed and be used for specific applications. Architectural forms are particularly useful when a user would like to create a DTD that adheres to an industry-standard DTD, but would still like to use his or her own customized tag names.

Since each DTD can serve, in principle, as a meta-DTD, architectural forms can be used to build up whole hierarchies of DTDs or standards. The property of such hierarchical DTDs to inherit the constructs of the meta-DTD a level above makes architectures very similar to classes in object-oriented systems. In fact, DTDs can even inherit from multiple enabling architectures.

The Kona Proposal for a Patient Record Architecture (PRA), an Enabling Architecture in Healthcare

Currently, attempts are being made to construct architectural forms for the healthcare sector by the Kona Editorial Group [18]. A three level architecture is being proposed as a Patient Record Architecture (PRA) for the exchange of clinical documents. The information model for this architecture is the HL7 Reference Information Model (RIM [19]). The least granular level of the architecture (Level 1) only encodes the header information, but leaves the rest of the document in plain text form. Level 2 structures the document into sections to allow minimal processing, whereas Level 3 will be consistent in granularity with the RIM. The idea of using an architectural form for structuring patient records is very promising, since it will hopefully allow the system to cope with the wide variety

of local needs for specific structures and still enable data exchange under the umbrella of a common standard.

Groves and Property Sets

The basic idea underlying groves is that SGML is merely a syntax for some underlying data model: the tree or, to be precise, a collection of trees. The grove provides a language for describing SGML's abstract data model. The grove is an abstract meta-"data-model" containing nodes with properties. The grove model for SGML itself or an SGML application like HyTime is specified with properties collected in *property sets*. The grove paradigm is defined in Annex A.4 to the HyTime standard (ISO/IEC 10744:1997) in the "Property Sets Definition Requirements (PSDR)" [20]. An introduction to groves and property sets by Paul Prescod can be found at [65].

Topic Navigation Maps (ISO/IEC 13250)

Topic Maps are an application of HyTime that uses its powerful model of universal addressing and independent linking. This international standard defines a notation for representing information about the structure of resources used to define topics, and the relationships between topics. Filters can be used either to include or to exclude information. A set of interrelated documents that employs the notation defined by this standard is called a Topic Navigation Map (TNM). Topic Maps are expressed as a set of architectural forms. As with HyTime, the TNM standard does not require a particular DTD to be used. It is an architecture that serves as template for adding attributes to elements in any DTD that can fit a specific environment [21]. There are no tools available at present to handle Topic Maps, but it is foreseeable that Topic Maps might play a significant role in the future as a syntax used to express semantic relations. Hopefully, such approaches will also be undertaken in the medical area where the problem of sementic diversity is considerable.

Document Style Semantics and Specification Language (DSSSL, ISO/IEC 10179:1996)

As discussed above, one of the basic principles of SGML is the separation of content and structure of a document from the style information used for its rendering in an output device. Complimentary to SGML, a standard was developed that defines rules for processing SGML documents: the Document Style Semantics and Specification Language (DSSSL). It contains three parts, a Style Language for formatting, a Transformation Language for transforming, and a Query Language for extracting data. DSSSL is closely related to SGML but is not an application of SGML. Information about DSSSL can be found at a site maintained by James Clark, the author of Jade, a freely available DSSSL-engine [22].

Back to the roots: XML - a subset of SGML

The success of HTML with the explosive expansion of the WWW is partly due to its relatively simple specifications, which allowed its rapid adoption by programmers and users on the Internet. However, due to its fixed DTD, each introduction of new element types required a new version of HTML, which made it somewhat inflexible. Indeed, software companies tended to extend HTML beyond its specifications, endangering its function as a quasi-standard. In 1996, a working group was

XSL•FO

RenderX

formed under the auspices of the W3C with the goal of solving the problem of extensibility. Remembering the roots of HTML, it considered adapting SGML with its open paradigm - which was so far not successful on the Web - to the needs of the Web community. The result of this endeavor was the creation of a subset of SGML, the eXtensible Markup Language (XML), that was easier for programmers to handle; this consequently encouraged widespread adoption.

XML 1.0 was accepted as a recommendation by the W3C in February 1998 and it has enjoyed an acceptance by the software industry that goes beyond all expectations. The specifications of XML have been annotated by one of its creators, Tim Bray, and can be found at [23]. For a more detailed introduction to XML see [24,25].

SGML/XML Applications

Even before its final specifications were accepted in 1998, several applications of XML were created by the community of users: Chemical Markup Language (CML) to manage chemical information [26]; or Mathematical Markup Language (MathML), for describing mathematical notation [27], which was accepted as a W3C Recommendation in April 1998 (revised July 1999).

Another early application of XML is the Resource Description Framework (RDF) for the processing of metadata and the provision of interoperability between applications that exchange machine-interpretable information on the Web [28]. It was recommended by W3C in February 1999.

An excellent compilation of SGML/XML applications is maintained by Robin Cover at OASIS's SGML/XML Web page under the headings:

- General Applications [29];
- Academic Applications [30];
- Industry Applications [31].

As an example we will discuss some applications in the healthcare sector.

XML in Healthcare

Healthcare is to a large extent an information-processing activity. Data about the patient's physical condition are collected by the treating physician using various diagnostic techniques, and evaluated within the framework of his or her medical knowledge to reach the appropriate decision for therapeutic measures or further diagnostic procedures. If this information processing path is to be effectively enhanced by electronic decision support systems, it is inevitable that data are to be structured at some time point, ideally at the very moment of data collection.

For this structuring to be useful, however, it requires a standard syntax and terminology that is used by all participating healthcare providers. The lack of such a commonly agreed-upon electronic language has so far been a major impediment for rapid development in this field. EDI (Electronic Data Interchange) standards like HL7 or UN/EDIFACT have found a certain application, but mainly in the administrative and financial areas of healthcare. For the first time, XML provides a concept and technology that promises to provide a flexible, open, and standardized solution to the problems of structuring, storing, and exchanging patient data. The independence of XML from particular software vendors, its self-describing nature, and not least, the fact that XML can be read by human beings as well as by computer programs, makes XML particularly suited to storing and handling documents and data over a long period of time, as it is needed for patient records.

HL7 SGML/XML Special Interest Group and the Task Force XML of CEN TC251

Health Level 7 (HL7) was founded in 1987 to develop standards for the electronic interchange of clinical, financial, and administrative information among independent healthcare oriented computer systems; e.g., hospital information systems, clinical laboratory systems, enterprise systems, and pharmacy systems. In August 1996, the HL7 Technical Steering Committee authorized the creation of an SGML Special Interest Group as part of a larger initiative to integrate SGML into medical informatics standards. "HCML" is a proposed abbreviation for the evolving markup language: "Health Care Markup Language" [32]. In December 1998, a draft document was produced as a proposal for using "XML as an Interchange Format for HL7 V2.3 Messages" [33]. In addition, an XML-based Patient Record Architecture (PRA) is being completed at Level One (see above).

In Europe, an XML task force has been established by CEN/TC 251 to investigate various aspects of using XML syntax for health messages and documents [34].

Figure 2. The SHCS Web document type definition and XML document (excerpts)

```
<!DOCTYPE SHCS [
...
<!ELEMENT LAB (REGULARTESTS,...)>
<!ELEMENT REGULARTESTS (...SUBPOPULATIONS,VIRALLOAD...)>
...
<!ELEMENT SUBPOPULATIONS  (DATE , LY-ABS , CD3-ABS , CD4-ABS , CD8-ABS ,
LY-REL , CD3-REL , CD4-REL , CD8-REL )>
<!ELEMENT LY-ABS   (#PCDATA )>
<!ELEMENT CD3-ABS  (#PCDATA )>
<!ELEMENT CD4-ABS  (#PCDATA )>
<!ELEMENT CD8-ABS  (#PCDATA )>
<!ELEMENT LY-REL   (#PCDATA )>
<!ELEMENT CD3-REL  (#PCDATA )>
<!ELEMENT CD4-REL  (#PCDATA )>
<!ELEMENT CD8-REL  (#PCDATA )>
...
<!ELEMENT DATE   ( (YYYY | MM | DD )* )>
<!ELEMENT DD  (#PCDATA )>
<!ELEMENT MM  (#PCDATA )>
<!ELEMENT YYYY  (#PCDATA )>
...
<!ELEMENT VIRALLOAD  (RNA , LIMIT , METHOD )>
<!ELEMENT RNA  (#PCDATA )>
<!ELEMENT LIMIT  (#PCDATA )>
<!ELEMENT METHOD  (#PCDATA )>
...
]>
<SHCS>
        ...
        <LAB>
             <REGULARTESTS>
                  ...
                  <SUBPOPULATIONS>
                       <DATE>
                            <YYYY>1966</YYYY>
                            <MM>12</MM>
                            <DD>16</DD>
                       </DATE>
                       <LY-ABS>1020</LY-ABS>
                       <CD3-ABS>800</CD3-ABS>
                       <CD4-ABS>360</CD4-ABS>
                       <CD8-ABS>460</CD8-ABS>
                       <LY-REL>34</LY-REL>
                       <CD3-REL>78</CD3-REL>
                       <CD4-REL>35</CD4-REL>
                       <CD8-REL>45</CD8-REL>
                  </SUBPOPULATIONS>
                  <VIRALLOAD>
                       <RNA>588</RNA>
                       <LIMIT>10</LIMIT>
                       <METHOD>sensitiv</METHOD>
                  </VIRALLOAD>
                  ...
             </REGULARTESTS>
             ...
        </LAB>
        ...
</SHCS>
```

SHCS Web: The use of XML in multi-center clinical studies

The extensibility of XML makes it particularly useful for the definition of syntactic rules and semantic conventions for communication within a domain of users with a specific task. We are using XML as WWW-based middleware in order to establish communication in a distributed and heterogeneous systems environment of a clinical multi-center study, the Swiss HIV Cohort Study (SHCS) [35].

Unlike HTML, XML allows for the explicit declaration of element types and representation of document structure in the DTD (see Figure 2). As stated for SGML, the names of the tags (i.e. the semantics of the terminal syntactical constructs) have to be fixed by convention. In our case, we use selected concepts as part of a domain-specific language (i. e. clinical immunology) as tag names (e.g. <viralload>). This approach yields the advantage that the XML documents can be interpreted by both machines and humans, thereby providing an excellent basis for cooperation among human and artificial agents.

The DTD for the SHCS Web application was set up using the preexisting paper-based study form as a template. So in essence, the paper study form was transformed into a structured electronic study form (ESF). One of the design problems in defining the DTD for the ESF was deciding in which cases data should be represented as attribute values and in which cases as element contents. As we began to work with XML even before February 1998, we faced the lack of others' experience with the standard. Thus, we decided in the beginning to pursue an attribute-oriented approach. In the meantime, the situation has changed. Today, a considerable number of XML applications exist, and practices, such as the previously mentioned question of how to represent contents, are being established. Thus, it appears that data are usually represented as element contents whereas attributes are preferably used for the representation of meta-information and references.

This change of DTD confronted us with the problem that we had to find a way to transform the old XML files to ones that followed the new DTD. We reached a solution by making use of the newly developed XSLT specification (see below) that provides an elegant way to deal with such problems. This course of events gives a perfect example of the flexibility of XML and the possibility of mediating between two different ad hoc standards of XML applications.

For more information about SHCS Web see Fierz & Grütter [36-38].

Other Examples of Healthcare Applications

The idea of using a generalized language for platform-independent structured reporting in healthcare had already been realized before the time of XML. A Data-entry and Reporting Markup Language (DRML), an SGML application, had been formulated by Kahn [39]. In Wales, a NHS project with the aim of structuring patient records had

originally been based on SGML, but has moved now to XML. Over 250,000 patient records from the Orthopaedic Hospital Trust in Oswestry have already been translated into XML [40]. The main advantages as seen by the initiators of the project lie in the possiblity of querying the patient's XML database in a much more efficient way than would have been possible with a free text search.

## *Doing Business with XML*

XML/EDI

According to a founding member of the XML/EDI Group initiative [41], the overall idea of XML/EDI is to add enough intelligence to the documents so that they become the framework for electronic commerce. Thereby, XML/EDI should define a standard for encoding the presentation characteristics, structure, and behavior of data that supports business transactions. Not only data should be delivered, but also information and the processing logic that is required to make sense of it. (Information is distinguished from data by its high-level semantics, referring to real-world objects, whereas the low-level semantics of data, i. e. the elementary data types, indicate only whether a given bit stream should be interpreted as CHAR, INTEGER etc.).

Webber identifies five components of an Integrated XML/EDI Internet-based System (see Figure 3):

- XML: The XML container transports the other components across the network. Thereby, XML tokens replace or supplement existing EDI segment identifiers. The system further takes advantage of the rich capabilities and transport layers of the Web and the Internet.
- EDI: Old EDI is called the grandfather of the current electronic commerce. Implementations include ANSI X.12 in the United States and UN/EDIFACT in Europe. XML/EDI provides 100% backward compatibility with existing EDI transactions, thereby preserving investments in existing EDI systems and knowledge.
- Templates: Process Templates enable processing of transactions (whereas Document Type Definitions (DTDs) enable transaction interoperability by defining the structure and content). They are globally referenced or travel along inside XML. Process Templates resemble traditional process control language syntax.
- Agents: Software Agents interpret the Process Templates to perform the work needed. They further interact with the EDI transaction data definitions and the users' business applications to create new templates for each new specific task.
- Repositories: Global Internet Repositories provide the semantic foundation for global business transactions. They allow for automatic lookups of the meaning and definition of the EDI elements. Global Internet Repositories are supplemented by adding DTDs and Process Templates. The XML/EDI group has compiled a "White Paper on XML Global Repositories for XML/EDI" [42].

Figure 3. An integrated XML/EDI Internet-based system (Webber, 1998, reproduction with permission from the author)
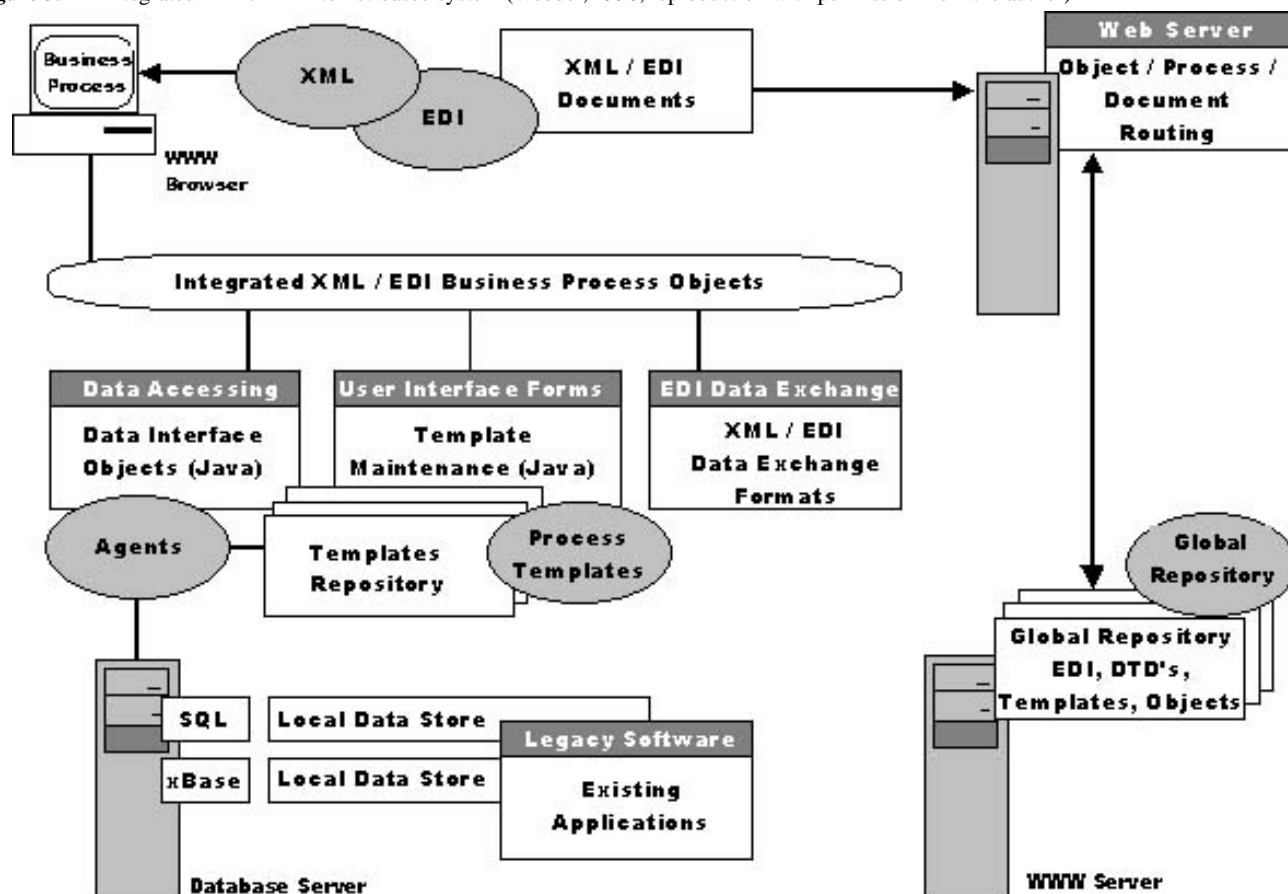
The real leverage of XML/EDI, when compared with traditional EDI, derives from the possibility of Partner A to communicate with Partner B based on the data format of their local business applications, instead of the necessity for both to conform to a standard. The system creates, with some manual help, a template describing the local record structures and field definitions. This template is added to the XML container and sent to Partner B along with the data. There it allows the receiving Partner B to create a second template defining the mappings of the data from Partner A onto the database system of his or her own business application as well as the rules that correspond to a set of necessary data transformations. During these processes, partner B is assisted by software agents. The first time, this mapping must be done manually. Once the template B is generated, data exchange from A to B can be automated. In other words, XML/EDI allows the creation of an ad hoc communication standard between A and B.

Currently, the XML/EDI framework is based on "Guidelines for using XML for Electronic Data Interchange" developed by the XML/EDI group [43]. XML/EDI pilot projects are under way for X.12-based data exchange [44], as well as for the European part of EDI, i.e., EDIFACT [45].

XMI

XML Metadata Interchange Format (XMI) is a new open industry standard that combines the benefits of the Web-based XML standard for defining, validating, and sharing document formats on the web with the benefits of the object-oriented Unified Modeling Language (UML). It provides application developers with a common language for specifying, visualizing, constructing, and documenting distributed objects and business models.

The objective of XMI is to allow the exchange of objects from the Object Management Group's (OMG) Object Analysis and Design Facility. These objects are more commonly described as UML and MOF (Meta Objects Facility) [46].

XML-related Specifications

Document Object Model (DOM)

Version 1.0, W3C Recommendation 1 October, 1998

> This specification defines the Document Object Model Level 1, a platform- and language-neutral interface that allows programs and scripts to dynamically access and update the content, structure and style of documents. The Document Object Model provides a standard set of objects for representing HTML and XML documents, a standard model of how these objects can be combined, and a standard interface for accessing and manipulating them. Vendors can support the DOM as an interface to their proprietary data structures and APIs, and content authors can write to the standard DOM interfaces rather than product-specific APIs, thus increasing interoperability on the Web [47].

An extension of the DOM Level 1 was being recommended by W3C in October 1999 as DOM Level 2 [48].

Namespaces in XML

To use DTDs and XML documents in a modular way, it is often desirable to be able to combine different DTDs within one XML document or to integrate XML documents that use different DTDs. However, since different DTDs might not recognize each other, they might use the same element names for different semantic entities and different element names for the same semantic entities, thereby inducing name collisions and semantic heterogeneity. In a W3C Recommendation of January 1999, rules for namespaces are described to solve this problem. XML namespaces provide a simple method for qualifying element and attribute names used in XML documents by associating them with namespaces identified by URI references [49].

Current and Future Developments

XML is just at its beginning. Its inclusion in a standardization framework paves the way for the development of further standards. Some of them are foreseeable and actually already under construction (see below). Others, however, might not yet be accessible to our imagination.

XSL

Extensible Stylesheet Language (XSL), an application of XML, is a language for expressing rendering information. It consists of a language for transforming XML documents, and an XML vocabulary for specifying formatting semantics. Thus XSL, like DSSSL that guided its conception, goes beyond merely specifying a syntax for defining style information. In a broader view, it is a transforming language that allows the conversion of documents obeying one DTD into documents referring to another DTD. It also contains elements for querying, what makes it a basis whereupon a query language can be built.

In March 2000, the W3C made the latest working draft of XSL 1.0 available [50]. In November 1999, the transforming part of XSL was issued as separate Recommendation called XSLT [51], and XPath designated as a language for addressing parts of an XML document, designed to be used by both XSLT and Xpointer (see below) [52]. A way to associate Style Sheets with XML documents has been specified and recommended by the W3C in June 1999 [53].

XQL

Beyond the original goal of SGML to standardize interchange of documents, XML will play an important role for interchange of any kind of data on the web. In fact, database tools that serve the XML standard are already on the market. The object-oriented paradigm is particularly suited for this purpose, but more importantly, a whole new approach in database design has already led to the first native XML database that preserves the original XML structure. However, a new XML-oriented query language is needed for searching, filtering, and retrieving data. In December 1998, the W3C convened a workshop on query languages called QL'98 [54]. A total of 66 position papers [55] from about 30 companies and 25 academic institutions and research facilities show the wide interest in the query issue. An extensive report on QL'98 has been compiled by Lisa Rein [56].

XLL

The success of HTML is in large part grounded on its simple linking mechanism that allowed programmers to turn the hypertext paradigm into worldwide reality. Preserving the linking functionality is therefore very important for XML. However, the linking mechanism provided by HTML again is somewhat limited, leading for example to the well-known problem of lost links on the WWW. The key concepts for an extended linking functionality are defined in the HyTime standard (see above). One basic idea is to separate the linking part from the addressing part to ease the maintenance of links. So, eXtensible Linking Language (XLL) as a broad term for XML hyperlinking (linking and addressing) has two major components: A linking language (XLink) and an addressing language (XPointer). XLink and Xpointer are currently (as of July 2000) candidate recommendations of the W3C [57,58].

XML Schema

The syntax of the Document Type Definition (DTD) is somewhat limited in its capability to express constraints on specific classes of documents. So, it does not provide, for example, a mechanism to describe primitive datatypes or default values for element contents. In February 1999, the W3C has issued requirements for a schema language [59] that uses XML syntax. Such a construct should allow, among other things, the import and export of datatypes from and to database systems, and the creation of user-defined datatypes. Two proposals for such a language have been submitted to the W3C: XML-Data [60], and Document Definition Markup Language (DDML) [61]. In April 2000, the W3C has issued two working drafts: XML Schema Part 1 for Structures [62], and XML Schema Part 2 for Datatypes [63].

XHTML

In January 2000 the W3C issued a recommendation for a reformulation of HTML 4.0 in XML 1.0: the Extensible HyperText Markup Language (XHTM™ 1.0). This specification defines HTML 4.0 as an XML application. The semantics of the elements and their attributes are defined in the W3C Recommendation for HTML 4.0. These semantics provide the foundation for future extensibility of XHTML. Compatibility with existing HTML user agents is possible by following a small set of guidelines [64]..

Conclusion

The described introduction of XML as a syntactical specification reflects a standardization process which is neither exclusively based on a binding decision of an acknowledged standardization authority (such as ISO) nor a pure market standard (or de facto standard) which is often a result of some monopolistic power (e.g., the operating system Windows). Instead, based on a standardization framework, a consortium of companies, academic institutions, and public bodies has agreed on a common recommendation. This story of the evolution of a standardization framework doubtlessly will end successfully in the case of XML, and we suggest that it should be considered as a generic model for standardization processes in the future.

The healthcare area is especially in need of such a standardization process, because of two main reasons: First, patient care has become more and more a process involving multiple providers, and rapid information exchange between the providers is pivotal not only to the patient's health, but also to the economic viability of healthcare. Second, the handling of patient data is a lifelong process that should not be affected by the ripples of vendor-specific software specifications.

With respect to the handling of semantic specifications, the old controversy of global standards that may be implemented as distributed and uniquely referable repositories versus local conventions has been adopted to the subject of semantic DTD schema specification. As mentioned, we think that the decision must not be in favor of either the global or local approaches, but should aim at their integration. There are actually two practical solutions to the problem of integration of local conventions into global standards. The first solution is a hierarchical approach which aims at aggregating local schemas into a general global schema. The second solution is a peer approach which aims at mediating between different schemas. The above-mentioned XML Stylesheet Language (XSL), particularly its transformations specification (XSLT), supports the second approach in that it allows a transformation between documents referring to different DTDs. Alternately, architectures allow for a mapping of different DTDs onto an aggregated one. However, as aggregation is always an abstraction from detailed information, the mapping of DTDs based on architectures can bring along a loss of information.

Figure 4. The logical connections between SGML/XML and related standards (abbreviations explained in the text)
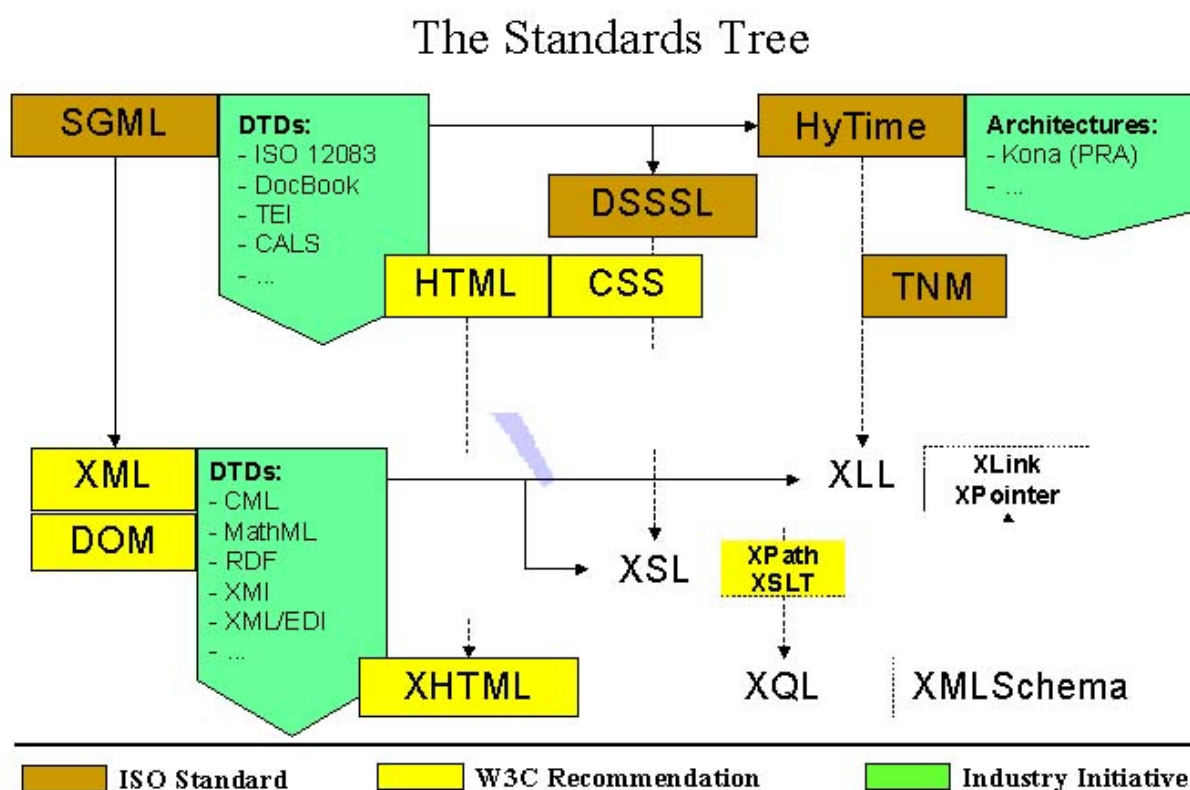
The possibility of constructing DTDs for transformation or architectural aggregation is an illustration of what we like to call a standardization framework (see Figure 4). But the story goes further: not only are SGML/XML meta-standards that have led to a rapid growth of applications in the form of new standards and meta-standards within just one year; but the process continues, particularly in the business area, and likewise in the healthcare domain. DTDs and templates will be established for particular business domains with an ever-increasing granularity of specification. It leads to a fractal behavior of the standardization system enabling a diversified growth within a common framework. This paradigm is close to nature.

Conflicts of Interest

None declared.

References

1.   World Wide Web Consortium (W3C) Members. URL: http://www.w3.org/Consortium/Member/List.html [accessed 2000 April 12]

XSL•FO
RenderX

2.   The Whirlwind Guide to SGML &XML Tools and Vendors. URL: http://tosca.infotek.no/sgmltool/guide.htm [accessed 2000 April 12]

3.   Alschuler L. HL7 document patient record architecture. HL7 12th Annual Plenary Session: XML and HL7. URL: http://www.hl7.org/library/committees/sgml/PRATutorial.ppt [accessed 2000 July 24]

4.   Alschuler L. ABCD...SGML: A User's Guide to Structured Information. London/Boston: Coriolis Group; Sep 1, 1995.

5.   Goldfarb CF. The SGML Handbook. Oxford, England: Oxford University Press; Feb 1, 1991.

6.   Steinmetz R. Multimedia-Technologie: Einführung und Grundlagen. Berlin; Heidelberg; New York: Springer; 1993.

7.   Megginson D. Structuring XML Documents. : Prentice Hall; Jan 15, 1998.

8.   The DocBook DTD. URL: http://www.oasis-open.org/docbook/ [accessed 2000 April 12]

9.   Text Encoding Initiative. URL: http://www.uic.edu/orgs/tei/ [accessed 2000 April 12]

10. CALS Standards. URL: http://navysgml.dt.navy.mil/cals/calsstds.html [accessed 2000 April 12]

11. Nelson TH. Computer Lib/Dream Machines. Redmond, Wash: Microsoft Press; Oct 1, 1987.

12. HTML 4.0 Specification. URL: http://www.w3.org/TR/REC-html40/ [accessed 2000 April 12]

13. Cascading Style Sheets, level 2, CSS2 Specification. URL: http://www.w3.org/TR/REC-CSS2/ [accessed 2000 April 12]

14. Information processing -- Hypermedia/Time-based Structuring Language (HyTime) - 2nd edition. URL: http://www.ornl.gov/sgml/wg8/docs/n1920/ [accessed 2000 April 12]

15. HyTime Users' Group Home Page. URL: http://www.hytime.org/ [accessed 2000 April 12]

16. Architectural Form Definition Requirements (AFDR). URL: http://www.ornl.gov/sgml/wg8/docs/n1920/html/clause-A.3.html [accessed 2000 April 12]

17. SGML ARCHITECTURES: Implications and Opportunities for Industry. URL: http://www.oasis-open.org/cover/newcomb-sgmlarch.html [accessed 2000 April 12]

18. HL7 Document. Patient Record Architecture. Framework Version 1.0: Draft. 2000 Apr 7. URL: http://www.hl7.org/special/committees/sgml/PRA/PRA_ballot_April00.zip [accessed 2000 July 24]

19. HL7 Reference Information Model. URL: http://www.hl7.org/Library/data-model/RIM/modelpage_non.htm [accessed 2000 July 24]

20. A.4 Property Set Definition Requirements (PSDR). URL: http://www.ornl.gov/sgml/wg8/docs/n1920/html/clause-A.4.html [accessed 2000 April 12]

21. ISO/IEC 13250, Topic Navigation Maps. URL: http://www.hightext.com/tnm/ [accessed 2000 April 12]

22. ISO/IEC 10179:1996. Document Style Semantics and Specification Language (DSSSL). URL: http://www.jclark.com/dsssl/ [accessed 2000 April 12]

23. Extensible Markup Language (XML) 1.0. URL: http://www.xml.com/axml/testaxml.htm [accessed 2000 April 12]

24. Bradley N. The Xml Companion. : Addison-Wesley (C); Sep 1, 1998.

25. Light R. Presenting Xml. Indianapolis, IN, United States of America: Sams; Sep 18, 1997.

26. XML-CML.ORG - The Site for Chemical Markup Language. URL: http://www.xml-cml.org/ [accessed 2000 April 12]

27. Mathematical Markup Language (MathML™) 1.01 Specification. URL: http://www.w3.org/TR/REC-MathML/ [accessed 2000 April 12]

28. Resource Description Framework (RDF) Model and Syntax Specification. URL: http://www.w3.org/TR/REC-rdf-syntax/ [accessed 2000 April 12]

29. SGML/XML Applications in Cross-Domain and Multi-Disciplinary Enterprises. URL: http://www.oasis-open.org/cover/gen-apps.html [accessed 2000 April 12]

30. SGML/XML: Academic Applications. URL: http://www.oasis-open.org/cover/acadapps.html [accessed 2000 April 12]

31. XML: Proposed Applications and Industry Initiatives. URL: http://www.oasis-open.org/cover/xml.html#applications [accessed 2000 April 12]

32. HL7 SGML/XML Special Interest Group X. URL: http://www.mcis.duke.edu/standards/HL7/sigs/sgml/ [accessed 2000 April 12]

33. XML as an Interchange Format for HL7 V2.3 Messages X. URL: http://www.mcis.duke.edu/standards/HL7/sigs/sgml/WhitePapers/hl7v2/hl7xml2.rtf [accessed 2000 April 12]

34. Task Force XML. URL: http://www.centc251.org/WItems/TFXml/Tfxml.htm [accessed 2000 April 12]

35. Ledergerber B, Von Overbeck J, Egger M, Lüthy R. The Swiss HIV Cohort Study: rationale, organization and selected baseline characteristics. Soz Praventivmed 1994;39(6):387-394. [Medline: 95117379] [doi: 10.1007/BF01299670]

36. SHCS Web Home. URL: http://www.knowledgemedia.org/knowledgemedia/knowledgemedia.nsf/pages/shcsweb_home.html [accessed 2000 April 12]

37. Fierz W, Grütter R. The use of a structured XML document to integrate distributed heterogeneous computer systems in a multi-centered clinical study. In: Greiser E, Wischnewsky M, editors. Methoden der Medizinischen Informatik, Biometrie und Epidemiologie in der modernen Informationsgesellschaft: Jahrestagung der GMDS. Bremen, München: MMV Medien & Medizin Verlag; 1998:324-326.

38. Grütter R, Stanoevska-Slabeva K, Fierz W. Enhancing knowledge management in a multi-center clinical trial by a web-based knowledge medium. In: Eder L, editor. Managing healthcare information systems with web-enabled technologies. Hershey, PA: Idea Group Publishing; 2000:124-140.

39. Kahn CE. A generalized language for platform-independent structured reporting. Methods Inf Med 1997 Aug;36(3):163-171. [Medline: 97439264]

40. SGML for Medical Records - The Oswestry Pilot Project. URL: http://www.keele.ac.uk/depts/rjah/sgml/sgml.htm [accessed 2000 April 12]

41. Webber DRR. Introducing XML/EDI frameworks. EM - International Journal of Electronic Markets 1998;8(1):38-41.

42. White Paper on XML Repositories for XML/EDI. URL: http://www.xmledi-group.org/xmledigroup/repository/xml-repWP.htm [accessed 2000 April 12]

43. Guidelines for using XML for Electronic Data Interchange. URL: http://www.geocities.com/WallStreet/Floor/5815/guide.htm [accessed 2000 July 24]

44. Joint CommerceNet GCA/X12 XML Workgroup. URL: http://www.commerce.net/projects/X12-XML/ [accessed 2000 April 12]

45. The European XML/EDI Pilot Project. URL: http://www.cenorm.be/isss/workshop/ec/xmledi/isss-xml.html [accessed 2000 April 12]

46. CORBA & XML Resource Page. URL: http://www.omg.org/technology/xml/index.htm [accessed 2000 July 24]

47. Document Object Model (DOM) Level 1 Specification. URL: http://www.w3.org/TR/DOM-Level-1/ [accessed 2000 April 12]

48. Document Object Model (DOM) Level 2 Specification. URL: http://www.w3.org/TR/WD-DOM-Level-2/ [accessed 2000 April 12]

49. Namespaces in XML. URL: http://www.w3.org/TR/REC-xml-names/ [accessed 2000 April 12]

50. Extensible Stylesheet Language (XSL) Specification. URL: http://www.w3.org/TR/WD-xsl [accessed 2000 April 12]

51. XML XSL Transformations (XSLT) Version 1.0. URL: http://www.w3.org/TR/xslt [accessed 2000 April 12]

52. XML Path Language (XPath) Version 1.0. URL: http://www.w3.org/TR/xpath [accessed 2000 April 12]

53. Associating Style Sheets with XML documents Version 1.0. URL: http://www.w3.org/TR/xml-stylesheet/ [accessed 2000 April 12]

54. QL'98 - The Query Languages Workshop. URL: http://www.w3.org/TandS/QL/QL98/ [accessed 2000 April 12]

55. QL'98 - Position Papers. URL: http://www.w3.org/TandS/QL/QL98/pp.html [accessed 2000 April 12]

56. The Quest for an XML Query Standard. URL: http://www.xml.com/xml/pub/1999/03/quest/index.html [accessed 2000 April 12]

57. XML Linking Language (XLink). URL: http://www.w3.org/TR/xlink [accessed 2000 April 12]

58. XML Pointer Language (XPointer). URL: http://www.w3.org/TR/xptr [accessed 2000 April 12]

59. XML Schema Requirements. URL: http://www.w3.org/TR/NOTE-xml-schema-req [accessed 2000 April 12]

60. XML-Data. URL: http://www.w3.org/TR/1998/NOTE-XML-data/ [accessed 2000 April 12]

61. Document Definition Markup Language (DDML) Specification, Version 1.0. URL: http://www.w3.org/TR/NOTE-ddml [accessed 2000 April 12]

62. XML Schema Part 1: Structures. URL: http://www.w3.org/TR/xmlschema-1/ [accessed 2000 April 12]

63. XML Schema Part 2: Datatypes. URL: http://www.w3.org/TR/xmlschema-2/ [accessed 2000 April 12]

64. XHTML™ 1.0: The Extensible HyperText Markup Language. URL: http://www.w3.org/TR/xhtml1/ [accessed 2000 April 12]

65. Prescod P. Addressing the Enterprise: Why the Web needs Groves. URL: http://www.prescod.net/groves/shorttut/ [accessed 2000 April 12]

e-Health Code of Ethics (May 24)

Helga Rippen, MD, PhD; Ahmad Risk, MD

Corresponding Authors:
Helga Rippen, MD, PhD
Internet Healthcare Coalition
USA
Email: hrippen@pfizer.com

Ahmad Risk, MD
Internet Healthcare Coalition
USA
Email: risk@thephiz.com

Abstract

The Internet is changing how people receive health information and health care. All who use the Internet for health-related purposes must join together to create an environment of trusted relationships to assure high quality information and services; protect privacy; and enhance the value of the Internet for both consumers and providers of health information, products, and services. The goal of the e-Health Code of Ethics is to ensure that people worldwide can confidently and with full understanding of known risks realise the potential of the Internet in managing their own health and the health of those in their care. The final e-Health Code of Ethics, presented in this paper, has been prepared as a result of the "e-Health Ethics Summit," which convened in Washington DC on 31 January 2000 - 2 February 2000. The summit, organized by the Internet Healthcare Coalition and hosted by the World Health Organisation/Pan-American Health Organisation (WHO/PAHO), was attended by a panel of about 50 invited experts from all over the world and produced the foundation for a draft code, which was released 18 February [1] for an online public consultation period which ended on 14 April 2000. The final Washington e-Health Code of Ethics sets forth guiding principles under eight main headings: candor; honesty; quality; informed consent; privacy; professionalism in online health care; responsible partnering; and accountability. Note: Abstract, keywords, acknowledgements and references have been added by the editor and are not part of the final Code.

Vision Statement

The goal of the **e-Health Code of Ethics** is to ensure that people worldwide can confidently and with full understanding of known risks realise the potential of the Internet in managing their own health and the health of those in their care.

Introduction

The Internet is changing how people give and receive health information and health care. All people who use the Internet for health-related purposes-patients, health care professionals and administrators, researchers, those who create or sell health products or services, and other stakeholders-must join together to create a safe environment and enhance the value of the Internet for meeting health care needs.

Because health information, products, and services have the potential both to improve health and to do harm, organisations and individuals that provide health information on the Internet have obligations to be trustworthy, provide high quality content, protect users' privacy, and adhere to standards of best practices for online commerce and online professional services in health care.

People who use Internet health sites and services share a responsibility to help assure the value and integrity of the health Internet by exercising judgment in using sites, products, and services, and by providing meaningful feedback about online health information, products, and services.

Definitions

Health information includes information for staying well, preventing and managing disease, and making other decisions related to health and health care.

- It includes information for making decisions about health products and health services.
- It may be in the form of data, text, audio, and/or video.
- It may involve enhancements through programming and interactivity.

XSL•FO
RenderX

Health products include drugs, medical devices, and other goods used to diagnose and treat illnesses or injuries or to maintain health. Health products include both drugs and medical devices subject to regulatory approval by agencies such as the U.S. Food and Drug Administration or U.K. Medicines Control Agency and vitamin, herbal, or other nutritional supplements and other products not subject to such regulatory oversight.

Health services include specific, personal medical care or advice; management of medical records; communication between health care providers and/or patients and health plans or insurers, or health care facilities regarding treatment decisions, claims, billing for services, etc.; and other services provided to support health care.

Health services also include listserves, bulletin boards, chat rooms, and other online venues for the exchange of health information.

Like health information, health services may be in the form of data, text, audio, and/or video, and may involve enhancements through programming and interactivity.

Anyone who uses the Internet for health-related reasons has a right to expect that organisations and individuals who provide health information, products or services online will uphold the following guiding principles:

Guiding Principles

Table 1. Guiding Principles

1. Disclose information that if known by consumers would likely affect consumers' understanding or use of the site or purchase or use of a product or service.	**Candor**
	
	People who use the Internet for health-related purposes need to be able to judge for themselves that the sites they visit and services they use are credible and trustworthy. Sites should clearly indicate
	• who owns or has a significant financial interest in the site or service
	• what the purpose of the site or service is
	  For example, whether it is solely educational, sells health products or services, or offers personal medical care or advice
	• any relationship (financial, professional, personal, or other) that a reasonable person would believe would likely influence his or her perception of the information, products, or services offered by the site
	  For example, if the site has commercial sponsors or partners, who those sponsors/partners are and whether they provide content for the site

2. Be truthful and not deceptive	**Honesty** People who seek health information on the Internet need to know that products or services are described truthfully and that information they receive is not presented in a misleading way. Sites should be forthright • in all content used to promote the sale of health products or services • in any claims about the efficacy, performance, or benefits of products or services They should clearly distinguish content intended to promote or sell a product, service, or organisation from educational or scientific content.

3. Provide health information that is accurate, easy to understand, and up to date.	**Quality**
	
	To make wise decisions about their health care, people need and have the right to expect that sites will provide accurate, well-supported information and products and services of high quality.
	
	To assure that the health information they provide is accurate, e-Health sites and services should make good faith efforts to
	• evaluate information rigorously and fairly, including information used to describe products or services
	• provide information that is consistent with the best available evidence
	• assure that when personalized medical care or advice is provided that care or advice is given by a qualified practitioner
	• indicate clearly whether information is based on scientific studies, expert consensus, or professional or personal experience or opinion
	• acknowledge that some issues are controversial and when that is the case make good faith efforts to present all reasonable sides in a fair and balanced way
	
	For example, advise users that there are alternative treatments for a particular health condition, such as surgery or radiation for prostate cancer
	
	Information and services must be easy for consumers to understand and use. Sites should present information and describe products or services
	• in language that is clear, easy to read, and appropriate for intended users
	  For example, in culturally appropriate ways in the primary language (or languages) of the site's expected audience
	• in a way that accommodates special needs users may have
	  For example, in large type or through audio channels for users whose vision is impaired
	
	Sites that provide information primarily for educational or scientific purposes should guarantee the independence of their editorial policy and practices by assuring that only the site's content editors determine editorial content and have the authority to reject advertising that they believe is inappropriate.
	
	Consumers have a right to expect that the information they receive is up to date. Sites should clearly indicate
	• when the site published the information it provides (and what version of the information users are seeing if it has been revised since it was first published)
	• when the site most recently reviewed the information
	• whether the site has made substantive changes in the information and if so, when the information was most recently updated

and	
Provide the information users need to make their own judgments about the health information, products, or services provided by the site.	Individuals need to be able to judge for themselves the quality of the health information they find on the Internet. Sites should describe clearly and accurately how content is developed for the site by telling users • what sources the site or content provider has used, with references or links to those sources • how the site evaluates content and what criteria are used to evaluate content, including on what basis the site decides to provide specific links to other sites or services For example, by describing the site's editorial board and policies When health products or services are subject to government regulation, sites should tell users whether those products (such as drugs or medical devices) have been approved by appropriate regulatory agencies, such as the U.S. Food and Drug Administration or U.K. Medicines Control Agency

XSL•FO
RenderX

4. Respect users' right to determine whether or how their personal data may be collected, used, or shared.	**Informed Consent**

People who use the Internet for health-related reasons have the right to be informed that personal data may be gathered, and to choose whether they will allow their personal data to be collected and whether they will allow it to be used or shared. And they have a right to be able to choose, consent, and control when and how they actively engage in a commercial relationship.

Sites should clearly disclose

- that there are potential risks to users' privacy on the Internet
  For example, that other organisations or individuals may be able to collect personal data when someone visits a site, without that site's knowledge; or that some jurisdictions (such as the European Union) protect privacy more stringently than others

Sites should not collect, use, or share personal data without the user's specific affirmative consent. To assure that users understand and make informed decisions about providing personal data, sites should indicate clearly and accurately

- what data is being collected when users visit the site
  For example, data about which parts of the site the user visited, or the user's name and email address, or specific data about the user's health or online purchases
- who is collecting that data
  For example, the site itself, or a third party
- how the site will use that data
  For example, to help the site provide better services to users, as part of a scientific study, or to provide personalised medical care or advice
- whether the site knowingly shares data with other organisations or individuals and if so, what data it shares
- which organisations or individuals the site shares data with and how it expects its affiliates to use that data
  For example, whether the site will share users' personal data with other organisations or individuals and for what purposes, and note when personal data will be shared with organizations or individuals in other countries
- obtain users affirmative consent to collect, use, or share personal data in the ways described
  For example, to collect and use the visitor's personal data in scientific research, or for commercial reasons such as sending information about new products or services to the user, or to share his or her personal data with other organisations or individuals
- what consequences there may be when a visitor refuses to give personal data
  For example, that the site may not be able to tailor the information it provides to the visitor's particular needs, or that the visitor may not have access to all areas of the site

"E-commerce" sites have an obligation to make clear to users when they are about to engage in a commercial transaction and to obtain users' specific affirmative consent to participate in that commercial transaction.

5. Respect the obligation to protect users' privacy.	**Privacy**

People who use the Internet for health-related reasons have the right to expect that personal data they provide will be kept confidential. Personal health data in particular may be very sensitive, and the consequences of inappropriate disclosure can be grave. To protect users, sites that collect personal data should

- take reasonable steps to prevent unauthorised access to or use of personal data
  For example, by "encrypting" data, protecting files with passwords, or using appropriate security software for all transactions involving users' personal medical or financial data
- make it easy for users to review personal data they have given and to update it or correct it when appropriate
- adopt reasonable mechanisms to trace how personal data is used
  For example, by using "audit trails" that show who viewed the data and when
- tell how the site stores users' personal data and for how long it stores that data
- assure that when personal data is "de-identified" (that is, when the user's name, email address, or other data that might identify him or her has been removed from the file) it cannot be linked back to the user

6. *Respect fundamental ethical obligations to patients and clients.*	**Professionalism in Online Health Care**

Physicians, nurses, pharmacists, therapists, and all other health care professionals who provide specific, personal medical care or advice online should

- abide by the ethical codes that govern their professions as practitioners in face-to-face relationships
- do no harm
- put patients' and clients' interests first
- protect patients' confidentiality
- clearly disclose any sponsorships, financial incentives, or other information that would likely affect the patient's or client's perception of professional's role or the services offered
- clearly disclose what fees, if any, will be charged for the online consultation and how payment for services is to be made
- obey the laws and regulations of relevant jurisdiction(s), including applicable laws governing professional licensing and prescribing

and

Inform and educate patients and clients about the limitations of on-line health care.	The Internet can be a powerful tool for helping to meet patients' health care needs, but users need to understand that it also has limitations. Health care professionals who practice on the Internet should clearly and accurately • identify themselves and tell patients or clients where they practice and what their professional credentials are • describe the terms and conditions of the particular online interaction For example, whether the health care professional will provide general advice about a particular health condition or will make specific recommendations and or referrals for the patient or client, or whether the health care professional can and will or cannot and will not prescribe medications in the particular situation • make good faith efforts to understand the patient's or client's particular circumstances and to help him or her identify health care resources that are available locally For example, to help the patient or client determine whether particular treatment is available in his or her home community or only from providers outside his or her community • give clear instructions for follow-up care when appropriate or necessary Health care professionals who offer personal medical services or advice online should • clearly and accurately describe the constraints of online diagnosis and treatment recommendations For example, providers should stress that because the online health care professional cannot examine the patient, it is important for patients to describe their health care needs as clearly they can • help "e-patients" understand when online consultation can and when it cannot and should not take the place of a face-to-face interaction with a health care provider
7. Ensure that organisations and sites with which they affiliate are trustworthy.	**Responsible Partnering** People need to be confident that organisations and individuals who operate on the Internet undertake to partner only with trustworthy individuals or organisations. Whether they are for-profit or nonprofit, sites should • make reasonable efforts to ensure that sponsors, partners, or other affiliates abide by applicable law and uphold the same ethical standards as the sites themselves • insist that current or prospective sponsors not influence the way search results are displayed for specific information on key words or topics And they should indicate clearly to users • whether links to other sites are provided for information only or are endorsements of those other sites • when they are leaving the site For example, by use of transition screens
8. Provide meaningful opportunity for users to give feedback to the site.	**Accountability** People need to be confident that organisations and individuals that provide health information, products, or services on the Internet take users' concerns seriously and that sites make good faith efforts to ensure that their practices are ethically sound. e-Health sites should • indicate clearly to users how they can contact the owner of the site or service and/or the party responsible for managing the site or service For example, how to contact specific manager(s) or customer service representatives with authority to address problems • provide easy-to-use tools for visitors to give feedback about the site and the quality of its information, products, or services • review complaints from users promptly and respond in a timely and appropriate manner Sites should encourage users to notify the site's manager(s) or customer service representatives if they believe that a site's commercial or noncommercial partners or affiliates, including sites to which links are provided, may violate law or ethical principles.
and Monitor their compliance with the e-Health Code of Ethics.	e-Health sites should describe their policies for self-monitoring clearly for users, and should encourage creative problem solving among site staff and affiliates.

Scott; Denise Silber; Bill Silberg; Loretta Skucas; S. Stuart Spector; Mike Squires; Harry Sweeney; David B. Vance; C. Peter Waegemann; Group Facilitators were Kathy Alison, Laura Guyer Miller, Maria Nagorski and Maureen Ponte.

Reference

1.   ; e-Health Ethics Initiative. e-Health Ethics Draft Code. J Med Internet Res 2000;2(1):e2 [FREE Full text] [doi: 10.2196/jmir.2.1.e2]

XSL·FO

RenderX

Letter to the Editor

Debate about evaluation and monitoring of sites carrying the HON-Logo

T Nater; C Boyer; G Eysenbach

Letter

Sir,

The JMIR editorial article of March 31, 2000 [1], made a number of careless and erroneous remarks about the HONcode (http://www.hon.ch/HONcode/). HON welcomes criticism as long as it's constructive, verified, based on fact, and fair. Your comments meet none of these criteria.

We want to rebut three points in particular.

1. The editorial says, "even quackery sites proudly display the [HONcode] logo" and provides a screenshot of http://www.selfhealthsolutions.com/breakthrough.htm. Your assertion is seriously exaggerated and the "evidence" you use is out-of-date. We know about this site. Last year, we demanded an end to its fraudulent and unauthorized use of the HONcode seal. The site complied and has not displayed it since.

All systems are vulnerable to abuse, but HON has a good record of quickly identifying and contacting most offenders. This is thanks to the sense of shared responsibility we encourage among Webmasters, information providers, and vigilant users of the health Internet. The author, too, should have alerted HON to the misuse of the HONcode as soon as he discovered it, instead of saving this cheap shot for publication.

2. You wrote that "the HON-Logo is a 'marketing trick,' to make the HONcode well known." We resent the belittling tone of this remark and deny any such thing. HONcode remains the medical Internet's first and most widely-supported ethical standard. HONcode membership is entirely free and HON makes no income of any kind from it. The HONcode has been translated from English into 12 foreign languages and has more than 3,000 member Web sites in 36 countries - hardly the track record of a "marketing trick."

3. You wrote "[the HONcode] is a toothless tiger. A more sophisticated system is needed, for example where the logo or 'seal of approval' is generated dynamically by a third party (as planned in the medCERTAIN project described below)."

You are stretching the reader's credulity by comparing the HONcode, a tried and trusted four-year-old product, with your future "project." And you ignore the facts. JMIR readers should know that the HONcode hyperlink seal is indeed generated dynamically. We run a policing system with proven effectiveness and are in proactive, executive control of the process at all times.

HON reviews all sites during the formal membership application process, insists on necessary changes and improvements, and periodically monitors them for continuing compliance thereafter. We list reviewed HONcode-compliant sites on our MedHunt search engine. HONcode membership always implies attribution to the relevant site of a unique, randomly-generated six-figure ID number, for example, http://www.hon.ch/HONcode/Conduct.html?HONConduct166259 for the National Library of Medicine's MEDLINEplus. The unique ID number, the Web site's URL, its HONcode registration status, correspondence and basic information on the site and its owner(s) are kept on a secure HON data base.

Finally, JMIR readers should be aware that we formally warn persistently non-compliant sites of the consequences of HONcode-related fraud and abuse. Our ultimate sanction is to de-link the HONcode seal on their pages. The vast majority of healthcare and medical Web sites prize their credibility and are striving to remain in business for the long term. This is why the threat of cut-off is usually enough to force problem sites either to comply with our demands or to remove our seal from their pages.

The HONcode will undergo a wide-ranging upgrade in the coming months to remain responsive to rapidly changing exigencies. While not a perfect system, it is the uncontested leader in its field. Too bad that the author of the JMIR article never sought to discuss with us some of the real challenges we face.

Timothy Nater, Executive Director

Celia Boyer, Manager of Web Services

Health On the Net Foundation, Geneva

References

1. Eysenbach G: Towards ethical guidelines for e-health: JMIR Theme Issue on eHealth Ethics. J Med Internet Res 2000, 2:e7. [Full text]

In reply

I regret that some of the remarks in the editorial [1] led to misunderstandings. As written in the editorial, the Health on the Net Foundation has been among the first to remind web publishers of their ethical duties, and is among the most successful initiatives in bringing these issues to a wide audience. Evidence-medicine has, however, taught us to stay critical; to continuously evaluate our interventions; and to ask how we can improve the effectiveness of our activities, especially if technology opens new possibilities. This letter gives me the

opportunity to clarify some of the issues touched on in the editorial and to hopefully eliminate any misunderstandings.

Referring to point 1, "even quackery sites proudly display the [HONcode] logo": The figure showing a questionable website bearing a HON logo was meant to illustrate the inherent vulnerability and limitations of a system that relies on a "second generation" technology, i.e. a static logo which can be simply copied and pasted by webmasters (as opposed to third generation technology, i.e. a dynamically generated logo or a client-side tool interpreting metadata - see below). It is unquestionable that HON can deal with cases of abuse, once they come to their attention. The point I was trying to make was to suggest measures that minimize the risk of such abuse occurring in the

first place, as opposed to a system that mainly relies on "post-marketing surveillance."

My proposal was that "the logo or seal of approval [could be] generated dynamically by a third party," suggesting a third generation of quality trustmarks, which can either be remotely loaded dynamic logos or PICS/XML/RDF-based metadata or both. In this approach, metadata and/or the logo would always come directly from the rating service to the user and could contain digital signatures, circumventing a necessary reliance on the co-operation of the health information provider (a number of sites carrying the HONCode-Logo have not actually included the dynamic backlink to HON).

Table 1. Definitions and Examples of Trustmarks

Generation	Characteristics	Examples
First generation trustmark	Logo/Award self-published by the information provider	HON up to 1999
Second generation trustmarks	Logo published by the information provider with hyperlink back to the rating service, generating a dynamic webpage	HON 1999, Etrust, Verisign
Third generation trustmarks	Metainformation coming directly from the rating service, either as PICS/XML/RDF metadata, or as dynamically generated seal	MedCERTAIN

The basic misunderstanding becomes apparent when the letter authors rebut this suggestion by saying that "you ignore the facts (...) the HONcode hyperlink seal is indeed generated dynamically." However, the HON seal (logo) itself is in fact *not* generated dynamically; only the page which will be generated if the user clicks the logo is dynamically generated. This is in fact a "second-generation" approach. A dynamic logo would be a logo which is for example remotely loaded from a third-part site, generated "on-the-fly," containing information

such as a timestamp, the URL for which it is valid, and information on the status of the site. Or it could even be generated at the client-computer of the user, if a special software could interpret metadata retrieved automatically from the provider and the evaluator. The example in Figure 1 demonstrates such a dynamically generated logo with a timestamp, generated at the University of Bristol (a MedCERTAIN partner) in real time.

Figure 1. Example of a third-generation trustmark, a dynamically (on-the-fly) generated logo, with timestamp for demonstration (note that this is not the actual MedCERTAIN trustmark, but only an illustration). The logo is actually remotely generated and loaded from the University of Bristol. Websites evaluated by MedCERTAIN will include a code on their website which remotely loads the logo from the MedCERTAIN website, and the logo ("trustmark") can contain "real-time" information

Referring to point 2: Regarding the irritation of Nater & Boyer concerning my quote "As the Health on the Net Foundation says, the HON-Logo is a 'marketing trick,' to make the HONcode well known," two issues should be pointed out:

First, I should re-emphasize that the expression "marketing trick" was a quote (and indicated as such) from an individual involved in HON at a conference in 1998. The quote was used to remind readers that the original idea of letting information providers publish the HON-Logo was to promote the code, and not to allow users to check the status of the site. Only recently has HON asked information providers to include a hyperlink back to HON with an ID, providing the user with the possibility of verifying the status of the site.

Giving out logos or awards to other websites with a backlink to the originating site is indeed a very common marketing "trick" on the web, and frankly referred to as such (see for example http://www.saltocompany.nl/marktng.html). It is a legitimate

marketing instrument to promote ideas, sites, products, or services. I doubt that the HONCode would have experienced a similar level of penetration if HON had relied on promoting the code in scholarly articles.

A second misunderstanding becomes apparent when the letter authors deny the idea of marketing and emphasize that HON "makes no income." The word "marketing" is commonly used in a nonprofit, public health context, without implying any financial purpose. Marketing has been defined (according to Kotler, quoted in [2], p. 29) as "human activity directed at satisfying needs and wants through exchange processes. An exchange process is simply the transfer, between two parties, of something that has value to each party. The marketer's task is to facilitate exchanges so that customers can fulfil their needs and wants. (...) Public sector and nonprofit marketers may benefit in nonmonetary ways, through the fulfilment of their institutional mission, desires, and goals." I see promoting the

quality of health information on the Internet as a new challenge for public health in the information age, and "public health practitioners are in the business of marketing [2]." With these definitions in mind, I cannot see what can be considered pejorative about the word "marketing."

Referring to point 3: I did *not* write "[the HONcode] is a toothless tiger," as erroneously quoted in the letter. The actual wording was "without third-party evaluation and enforcement (...), this ethical code is a toothless tiger." "*This* ethical code" was in fact primarily referring to the Washington Code of eHealth ethics, which was the main topic of the editorial. Only three sentences of the editorial actually referred to HON. A clearer formulation would perhaps have been "without third-party evaluation and enforcement, *any* ethical code is a toothless tiger." It is noteworthy that a few weeks after publication of the editorial, a member of the Hi-Ethics group (another US group trying to implement a code of ethics for health websites) used almost the same words as used in the editorial:

> *Members of the group agree that devising enforcement mechanisms is the necessary next step. (...) "What doesn't work is a code with no follow-up and no teeth," [said Mark Boulding, a member of the group, in the context of the Hi-Ethics code] [3].*

Neither this statement nor the statement in the editorial are meant to imply that either the HONcode or the Washington Code of eHealth ethics has no teeth, but only to point out that ethical codes have to be complemented by appropriate measures of evaluation and enforcement.

The question remains whether and to what extent sites carrying the HON logo are actually evaluated and monitored. The letter above suggests a rigorous evaluation, asserting that "most offenders" are "quickly identified." I would be interested in the actual percentage, i.e. how many "offenders" are identified per year, and what is the denominator? I.e. out of how many sites carrying the HON logo how many are not sticking to the HONcode? Shon & Musen [4] have recently shown that "HONoured" websites, found in HON's Medhunt, display authorship, currency of information, references, and disclosure of sponsors *less* often than a random selection of sites found in AltaVista. These data should be reason enough for HON to attempt to produce hard data on the actual quality of sites carrying the HON logo.

While the letter above suggest that sites registering at HON are evaluated in some way ("HON reviews all sites during the formal membership application process"), HON makes contradicting statements on their website (http://www.hon.ch/HONcode/), saying that "the HONcode is not an award system, nor does it intend to rate the quality of the information provided by a Web site. It only defines a set of rules to hold Web site developers to basic ethical standards in the presentation of information and to help make sure readers always know the source and the purpose of the data they are reading." If HON does not rate quality, on what grounds are quackery sites such as the one shown in the editorial's illustration asked to remove the HON-logo?

My main concern here is that, provided with remaining vague and sometimes contradictory information on this issue, the public can be easily confused. Even in the peer-reviewed literature, HON is sometimes referred to as a organization which rates the quality of websites, and a number of sites publish the HON-Logo under the heading "Awards" (which implies evaluation). MEDLINEplus, for example, which is cited in the letter, displays the HON-Logo under the heading "Awards." HON does not seem to regard this as "HON related fraud and abuse" and takes no measures to avoid this impression, although HON itself says that it is not an award system. This issue should not be taken too lightly. Consumers may erroneously mistake the HON-Logo as an award and rely on it as an indicator for assessed information.

The notion expressed by the authors in the letter - that issues concerning quality assessment were never discussed with them - is different from my own recollection. On the contrary, during my last conversation with one of the letter authors in Washington in February 2000, I had the impression that HON realizes the boundaries of their current approach and is willing to work with MedCERTAIN to improve the current system.

MedCERTAIN should not be seen as a competition or threat to HON, as it has a clearly different focus. While it also encourages self-regulation, it is very clear about the fact that it will actually rate (evaluate) websites and that it will build a technical platform for the interoperability of existing rating and reviewing services.

I made clear in the editorial and elsewhere [5] that both approaches (encouragement of self-regulation by HON, and third-party evaluation by organizations collaborating with MedCERTAIN) are important pillars towards quality management of health information on the Web, alongside educating users and enforcing "best practice" codes. We see the activities of MedCERTAIN as complementary to self-regulation efforts such as HON and the Washington eHealth Code of Ethics. We think that HON could well benefit from the technical and organizational infrastructure that will be provided by MedCERTAIN, and have from the beginning invited HON to join this collaboration. The future of monitoring, assessing, and evaluating health information certainly does not lie in reliance on a single, central organization, but in a democratic, decentralized, distributed and collaborative system [6].

Gunther Eysenbach

JMIR Editor-in-chief

References

1. Eysenbach G. Towards ethical guidelines for e-health: JMIR theme issue on eHealth ethics. J Med Internet Res 2000 Mar 31;2(1):e7 [FREE Full text] [Medline: 21577991] [doi: 10.2196/jmir.2.1.e7]
2. Siegel M, Doner L. Marketing Public Health: Strategies to Promote Social Change. Gaithersburg, MD: Jones & Bartlett Publishers; Sep 15, 1998.
3. Schwartz J. Online Health Sites to Unveil Standards. Washington: Washington Post; 2000:A12.
4. Shon J, Musen MA. The low availability of metadata elements for evaluating the quality of medical information on the World Wide Web. Proc AMIA Symp 1999:945-949 [FREE Full text] [Medline: 20033010]
5. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [Medline: 20323060] [doi: 10.1136/bmj.320.7251.1713]
6. Eysenbach G, Diepgen TL. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ 1998 Nov 28;317(7171):1496-1500 [FREE Full text] [Medline: 99051215]

XSL•FO
RenderX

<u>Original Paper</u>

Organizing the Present, Looking to the Future: An Online Knowledge Repository to Facilitate Collaboration

Charles Burchill[1], BSc, MSc; Leslie L Roos[1], PhD; Patricia Fergusson[1], PhD; Laurel Jebamani[1], BA; Ken Turner[1], BSc; Stephen Dueck[1]

Manitoba Centre for Health Policy and Evaluation, Department of Community Health Sciences, Faculty of Medicine, University of Manitoba, St. Boniface General Hospital Research Centre, Canada

Corresponding Author:
Leslie L Roos, PhD
Manitoba Centre for Health Policy and Evaluation
St. Boniface General Hospital Research Centre
Rm R2008 - 351 Tache Avenue
Winnipeg Manitoba R2H 2A6
Canada
Phone: +1 204 235 3719
Fax: +1 204 237 6641
Email: Leslie_Roos@cpe.umanitoba.ca

Abstract

Background: Comprehensive data available in the Canadian province of Manitoba since 1970 have aided study of the interaction between population health, health care utilization, and structural features of the health care system. Given a complex linked database and many ongoing projects, better organization of available epidemiological, institutional, and technical information was needed.

Objective: The Manitoba Centre for Health Policy and Evaluation wished to develop a knowledge repository to handle data, document research methods, and facilitate both internal communication and collaboration with other sites.

Methods: This evolving knowledge repository consists of both public and internal (restricted access) pages on the World Wide Web (WWW). Information can be accessed using an indexed logical format or queried to allow entry at user-defined points. The main topics are: Concept Dictionary, Research Definitions, Meta-Index, and Glossary. The Concept Dictionary operationalizes concepts used in health research using administrative data, outlining the creation of complex variables. Research Definitions specify the codes for common surgical procedures, tests, and diagnoses. The Meta-Index organizes concepts and definitions according to the Medical Sub-Heading (MeSH) system developed by the National Library of Medicine. The Glossary facilitates navigation through the research terms and abbreviations in the knowledge repository. An Education Resources heading presents a web-based graduate course using substantial amounts of material in the Concept Dictionary, a lecture in the Epidemiology Supercourse, and material for Manitoba's Regional Health Authorities. Confidential information (including Data Dictionaries) is available on the Centre's internal website.

Results: Use of the public pages has increased dramatically since January 1998, with almost 6,000 page hits from 250 different hosts in May 1999. More recently, the number of page hits has averaged around 4,000 per month, while the number of unique hosts has climbed to around 400.

Conclusions: This knowledge repository promotes standardization and increases efficiency by placing concepts and associated programming in the Centre's collective memory. Collaboration and project management are facilitated.

XSL•FO
RenderX

Introduction

Worldwide, an increasing number of research centers are using administrative databases for monitoring and evaluating health care systems. In several Canadian provinces, comprehensive health care data are made available for research not only by an insurance plan that covers almost the entire population, but also by a strong commitment to developing this resource. The Manitoba Centre for Health Policy and Evaluation (MCHPE), a university-based research unit, has been using the Manitoba Health database since 1991. MCHPE's founders began working with these data in 1975; the database has been developing incrementally since 1970.

Figure 1. An Ideal Administrative Database

Figure 1. An Ideal Administrative Data Base
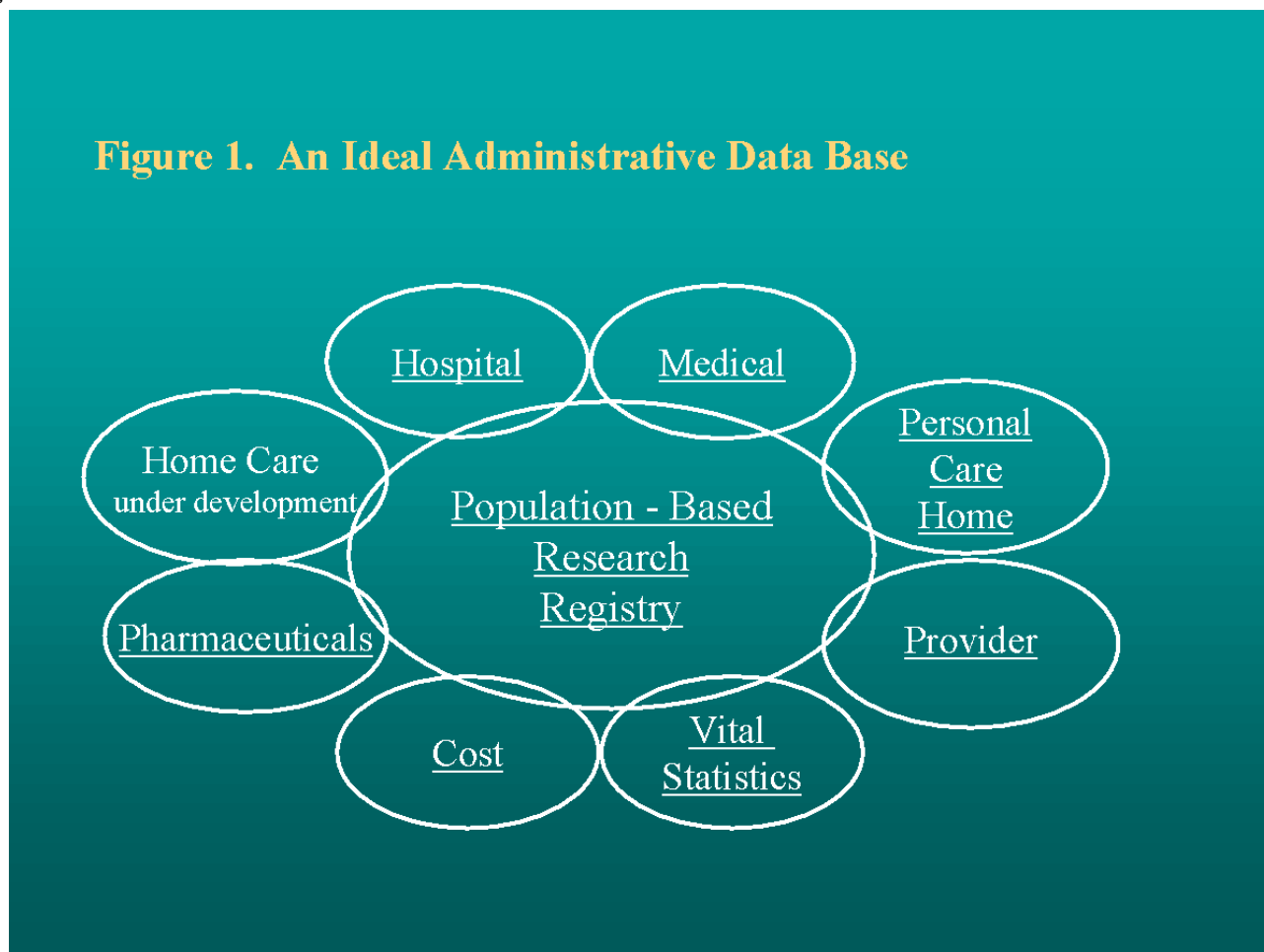
Manitoba's database documents almost every contact that the population has with the health care system. Figure 1 presents an idealized view of the MCHPE database with a population-based research registry playing a central role. The research registry contains a unique, encrypted number assigned to each provincial resident, together with information on demographic characteristics, location of residence, and family composition. The substantive files include information on hospital stays, physician visits, nursing home stays, pharmaceutical use, and so forth. These files use the same number to track utilization of services, making it possible to compile comprehensive histories for individuals over time. Similar databases outside Canada (Oxford Record Linkage Study, Scottish Record Linkage System, Rochester Epidemiology Project, Western Australian Health Services Research Linked Database) have recently been reviewed. The Manitoba database was noted as representing "international best practice" across a number of benchmarks [1]. Approximately 300 papers or reports have been written using the Manitoba data (MCHPE Papers Published).

With the advent of technologies facilitating linkages between databases, Manitoba researchers have been better able to study the complex relations between population health, health care utilization, and structural features of the health care system [2]. As part of its long-range plan, MCHPE hopes to expand its database by linking the existing health files to population-based data on students, families, and classrooms from Manitoba's Department of Education.
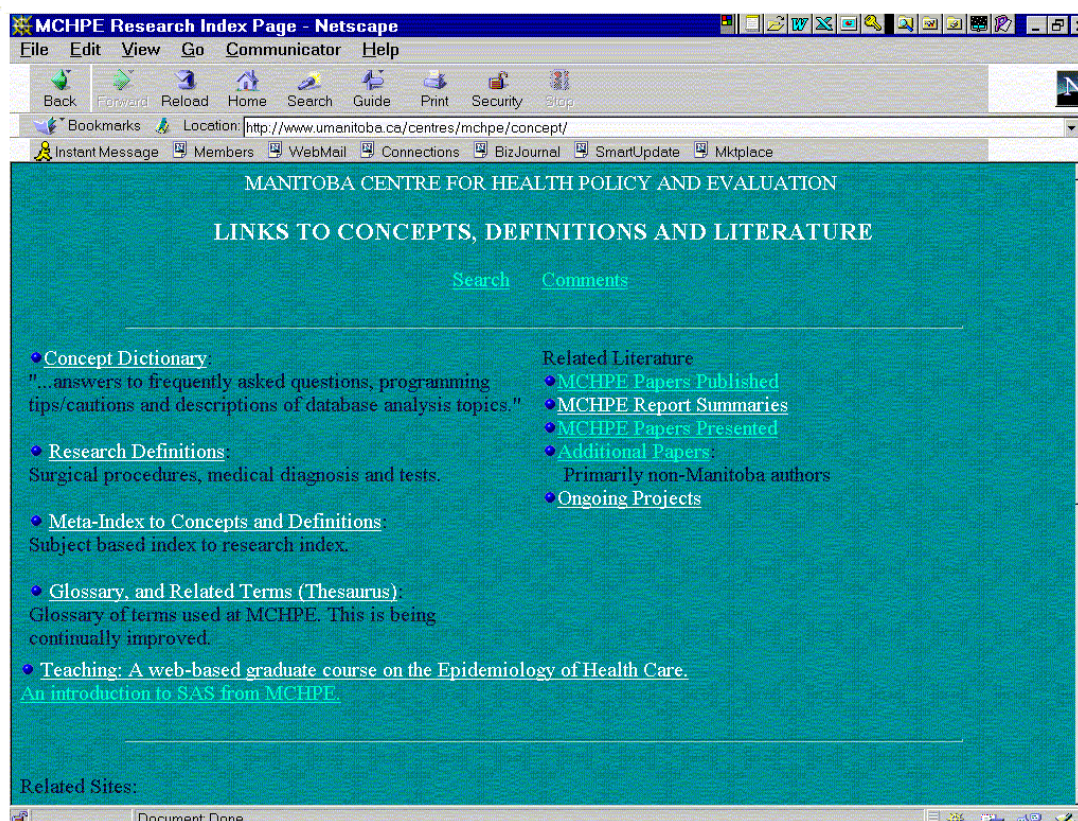
MCHPE has required an efficient system to manage this information resource, to document research methods, and to facilitate internal communication and collaboration with other sites. Within Winnipeg, MCHPE has staff members located at three different University of Manitoba campuses. Additional staff and active collaborators are distributed throughout North America. As much as possible, we wish to facilitate work and interaction as if the researchers were in the same physical space [3].

Over nine years of operation, MCHPE has amassed a vast amount of epidemiological, institutional, and technical

information. It can be a formidable task to identify the local expert to assist with a particular topic or to access up-to-date material on orientation; research projects and protocols; data dictionaries; and guidelines for security, access, and confidentiality. How do we update documentation and ensure consistency to permit replication and extension of studies? How do we prevent re-creating the wheel after other investigators have already operationalized research concepts? This paper describes the development, current status, and future directions of the "repository of shared knowledge" at MCHPE [3]. We have been guided not only by the need for efficient data management and standardized research methods within MCHPE, but also by the wish to make the "Centre's intellectual assets more widely available to scholars, students and other health care researchers" [4]. Our long-term goal is to produce a knowledge system that incorporates concepts and methods from other sources, documenting alternative perspectives and mentioning shortcomings or potential problems.

Methods

Knowledge Repository

The MCHPE knowledge repository is an evolving enterprise, consisting of both external (i.e., public) and internal Web pages [5]. External users can access the public website available over the Internet as part of the MCHPE home page on the World Wide Web (WWW) (URL: http://www.umanitoba.ca/centres/mchpe). The internal pages are accessible only by authorized individuals.

Four main headings (under the Concepts/Research button on the home page) organize information in the external website: Concept Dictionary, Research Definitions, Meta-Index to Concepts and Definitions, and Glossary and Related Terms (Thesaurus). Two other headings, Teaching and Related Literature, are designed to help researchers. Cross-links facilitate information retrieval. Figure 11 illustrates how the information available to external users is organized.

Figure 11. Research Index Page

Table 1: Research Index Page

The Concept Dictionary provides one (or more) operational definitions of concepts used by health researchers, and describes in detail the development of new variables or creation of variables based on existing data (e.g., age from birth year; Full Time Equivalent (FTE) Physicians from physician billing numbers, billing location, individual patient visitation). The creation of variables requiring a complex set of steps is also outlined. The content of the Concept Dictionary, like that of the Data Dictionaries (see below), was designed to provide the information necessary for research continuity.

Figure 12 presents a sample of the terms included in the Concept Dictionary. To illustrate, the entry under "Ambulatory Care Sensitive Conditions" defines the concept, cites relevant research from which the concept was originally defined, and directs those seeking additional information to the MCHPE programmer or researcher most familiar with the concept. The entry under "Age" describes various methods for calculating age and format

requirements. Entries often include cautions associated with the concept, as well as possible remedies. For example, the write-up for calculating the exact age for a particular date warns that Manitoba physician claims do not generally contain the "date of birth" field and suggests how this information can be acquired.

Some concept definitions call for a more lengthy discussion of measurement. Various issues associated with the Manitoba database are outlined, including the following: standard data exclusions, strategies for avoiding double-counting of hospital visits over multiple years and concurrent stays, and definitions of geographical regions. The entry "Health Status Indicators" includes a section describing six indicators, a section of related

definitions, a section noting key references, and a table listing various indicators and their associated ICD-9-CM codes.

The Concept Dictionary is organized alphabetically with numerous cross-references to facilitate logical connections among various dictionary entries. For example, MCHPE plans to develop linkages from concepts such as "income group calculation" to specific variables (e.g., municipal and postal codes) in the data dictionary from which it was derived. The "Research Definitions" heading (Figure 13) defines the codes for common surgical procedures, tests, and diagnoses derived from classification systems such as ICD-9-CM and physician tariff codes, and from common groupers such as DRG, RDRG, and CMG. Where relevant, cross linkages are made to Concept Dictionary entries.
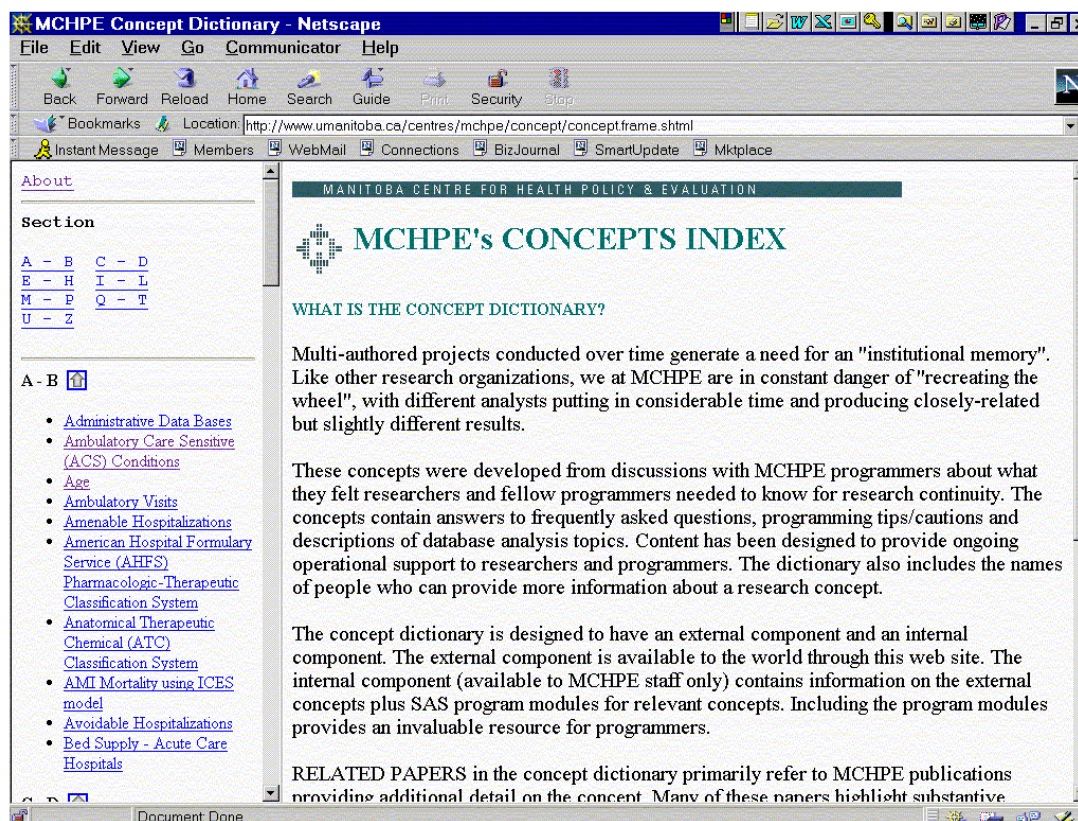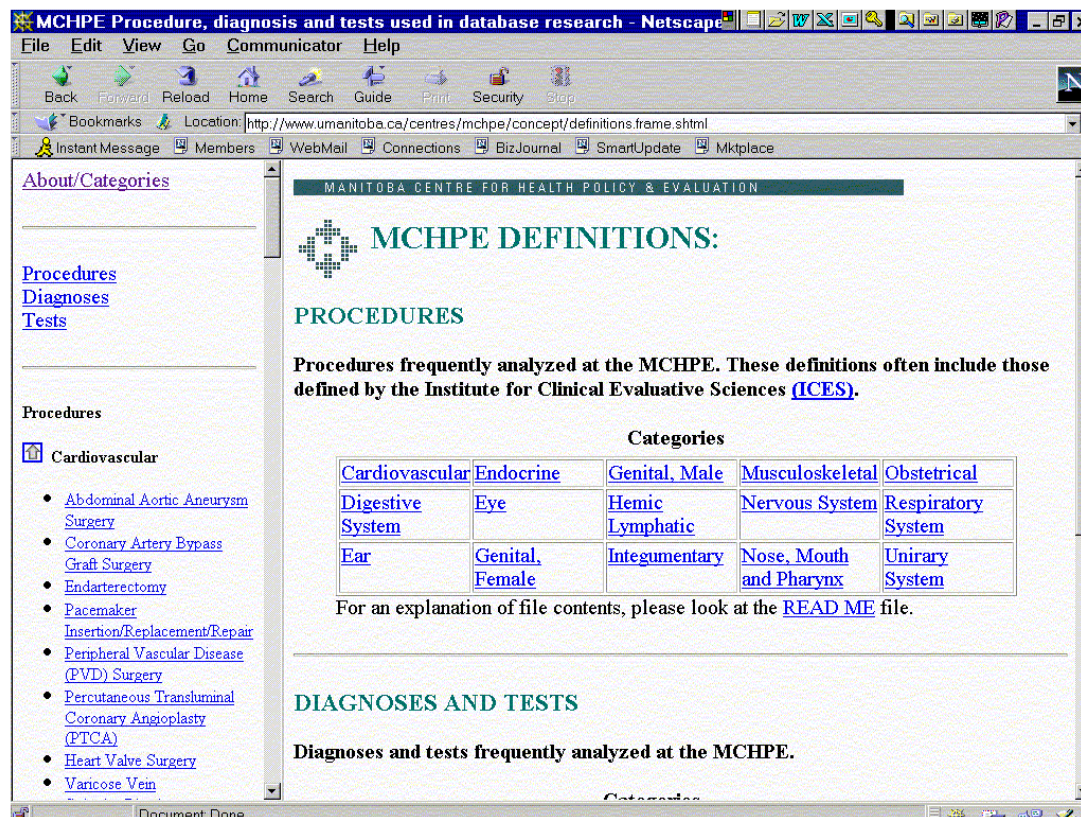
Figure 12. Concept Dictionary

Figure 13. Research Definitions

Table 3: Research Definitions

The Meta-Index to Concepts and Definitions is designed to make information retrieval easier. The Meta-Index organizes concepts and definitions according to the Medical Sub-Heading (MeSH) system developed by the National Library of Medicine. MeSH allows the researcher to start with a broad category of interest and increase the specificity by choosing pertinent "branches" of the Meta-Concept "tree" (Figure 2). Although MeSH does not currently contain any content regarding public health, necessary modifications were kept to a minimum to adhere as closely to the original standard as possible.

A Glossary of terms facilitates access to information by a broader audience. The range of research terms and abbreviations can make navigation through areas of the knowledge repository unwieldy. More than one thousand terms from MCHPE government deliverables and research papers were generated as possible entries. The Glossary is organized alphabetically and designed as a quick reference to possibly unfamiliar terms encountered while reading MCHPE work. Currently, the content under each entry in the Glossary includes a brief definition (as many as are relevant), a list of synonyms and related terms, a reference list for papers that use the concept, and, for more in-depth information, links to relevant parts of the Concept Dictionary (Figure 14).
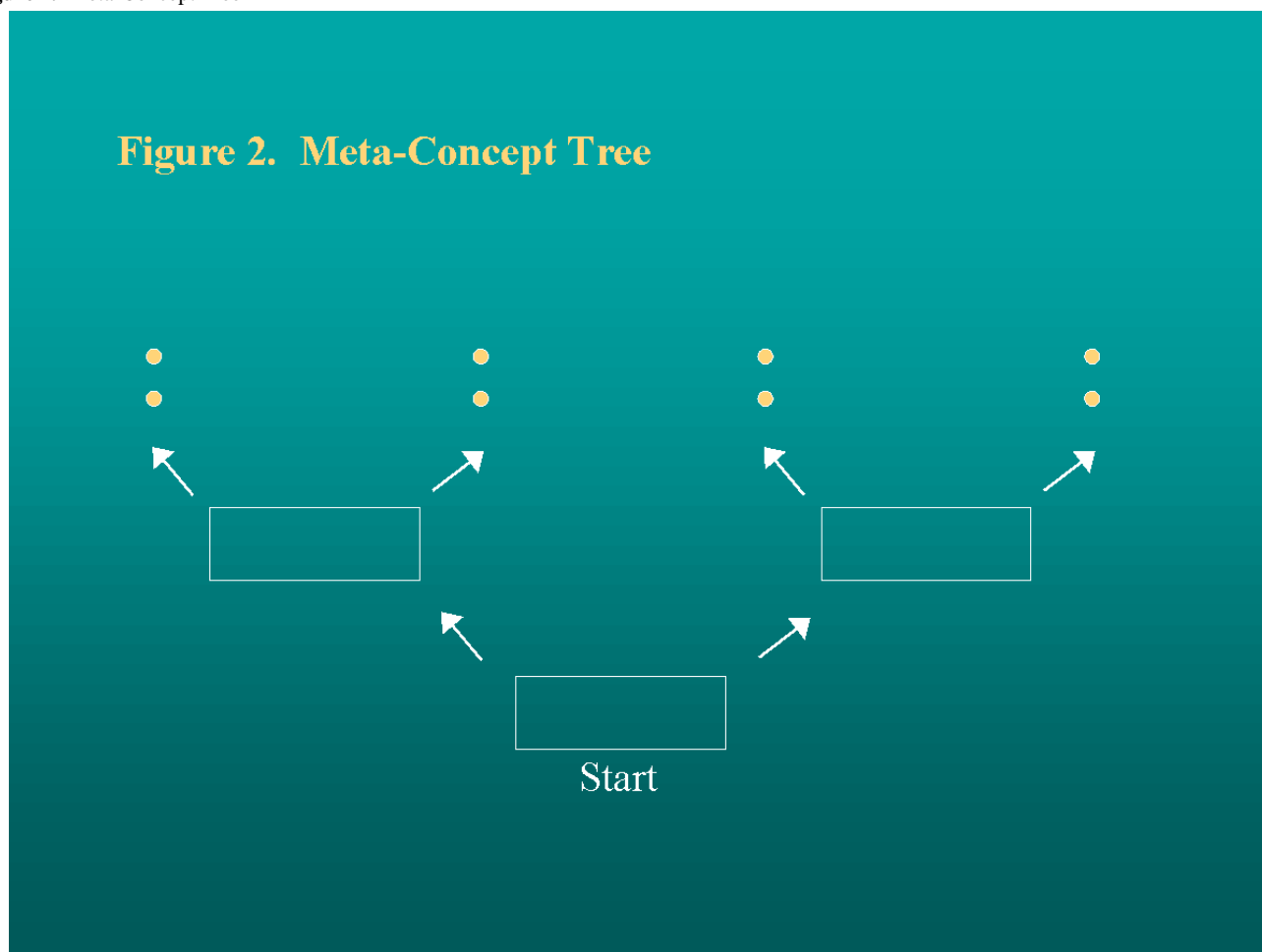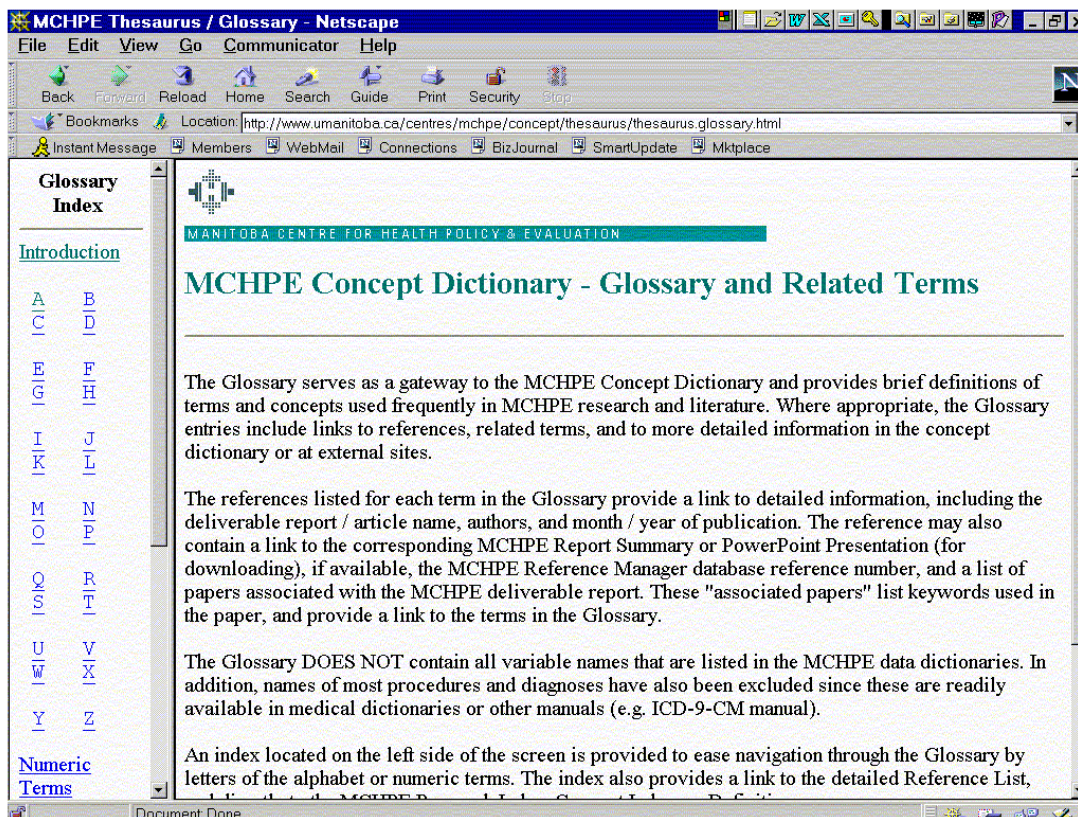
Figure 2. Meta-Concept Tree

Figure 2. Meta-Concept Tree

Figure 14. Glossary

Table 4: Glossary

Teaching presents the web-based version of a graduate course, "The Epidemiology of Health Care."" Given the typical course audience - a substantial number of auditors and a small number of enrolled graduate students - most of the weekly readings use links to material already available on the web: project summaries, newsletters, and entries in the Concept Dictionary. Many of the academic papers distributed in paper form are from MCHPE's June 1999 supplement to Medical Care. The Glossary incorporates links from the definition of terms to MCHPE deliverables and research papers where they appear. For teaching purposes, links that go in the other direction allow the student to produce a list of key concepts and terms associated with a given research paper or report. The course and other teaching materials (including a lecture in the Epidemiology Supercourse) are included under the Education/Resources button on the home page. Finally, the heading Related Literature includes: MCHPE Papers Published (with accompanying abstracts for recent years), MCHPE Report Summaries, MCHPE Papers Presented, Additional Papers (primarily non-Manitoba authors), and Ongoing Projects using the research definition(s).

Links to associated pages (both locally and elsewhere on the Internet) allow quick cross-references and navigation through complex topics. The search engine (Excite) allows users to search the Research Index using key words or phrases. Conversion of all documents available at the MCHPE website to HTML format, together with the addition of meta-tags with key words, has provided faster, more reliable searches. The Concept Dictionary has also been added to many general external search engines (Lycos, Hotbot, and Excite) to facilitate access.

Internal Website

Material that is restricted in nature or irrelevant to external users is available on MCHPE's internal website. Some concepts include SAS programming, and incorporate confidential information such as hospital identifier. In other cases, exclusion from the public website is based on the proprietary nature of the code or database structure. Specific file names or locations provide structural information regarding our computer systems; external users are also excluded from this information. Finally, use of the internal dictionaries assumes familiarity with MCHPE data organization; variable names meaningful to only MCHPE staff would not be useful on the external web site.

Data Dictionaries are highly valued among information systems users; they typically list and describe variables and the structure of specific files of any given set of databases. MCHPE has separate data dictionaries for each major set of data listed in Figure 1(e.g., hospital, medical, nursing home, provider, vital statistics, cost, pharmaceuticals). These data dictionaries are routinely stored as part of the various MCHPE data files. The contents include standard names for variables, formats, interpretation of the variables, and an overview of known problems and recommendations. Pages are designed to facilitate access to database research concepts and definitions. Given the confidential nature of the information, the data dictionaries are not currently available on the external Web pages.

Quality and Bias

Notwithstanding the wider applicability of MCHPE-produced work, our knowledge repository has emphasized ideas and methods associated with one research center. The risk of bias through presenting a single viewpoint needs to be addressed, particularly as the Concept Dictionary expands and perhaps becomes a "destination site" for health care researchers using administrative data. Unlike peer-reviewed papers, the content of web sites generally is not filtered through a critical academic review process. Online knowledge systems must develop and incorporate methods for ensuring information quality. Our procedures to address this issue might help others:

1. Develop clear standards and procedures for determining the content of the knowledge system. Clearly, quality is related to the process by which information is added; some items in the Concept Dictionary are identified as coming from papers and reports written elsewhere. Concepts originating in Manitoba must be tested with MCHPE data, for example, before being entered in the Concept Dictionary. The principal researcher or programmer typically works with an experienced research assistant to write entries for the Concept Dictionary as part of the documentation process for any research project or publication. Additionally, concepts are added using a style which notes:
   a. Name of concept and date of information presented on the page.
   b. A basic description, including both situations where the concept is relevant and limitations or cautions on use of the concept.
   c. SAS code (or location of such code) and/or SAS macros used to create the concept [6]. (At MCHPE almost all programming is performed using SAS. When other programming languages are used, the appropriate code is provided).
   d. SAS formats and necessary labels (or their location).
   e. Other information such as relevant contact(s) and any published papers using the concept.

2. Develop procedures to keep the knowledge system current and accurate. MCHPE is considering a position for a "knowledge expert," an individual with substantive knowledge of the databases and research concepts who can develop quality-control procedures.

3. Solicit feedback. Other users of the knowledge base may supplement the information provided or define a concept differently. At MCHPE, we are seeking feedback via an entry on the Concept Dictionary home page. Other possibilities for broadening users' perspectives include incorporating information on new concepts and alternative definitions, and increasing the number of links to other research centers.

4. Make evaluation an integral part of development and maintenance. Currently, MCHPE is documenting who is using different parts of the web site.

5. Be explicit. The process of making information explicit permits individuals to evaluate, use, or develop it in ways that ultimately advance the research process.

Results

The MCHPE website continues to reach an increasing number of national and international health administrators, researchers, and students. Reports, report summaries, and newsletters are available from the home page. In May 1999, MCHPE received close to 16,000 hits on the website's pages; almost 6,000 of these used the knowledge repository. Web traffic grew at an average rate of over 35% each month during the first part of 1999 (Figure 3 and Figure 4). Since then, traffic on the knowledge repository has dropped to around 4,000 page hits per month while the number of unique hosts has climbed to around 400. Depending upon whether or not a course is scheduled, between 1,000 and 3,000 hits per month have been recorded on the Education pages. When new reports are issued on the MCHPE web site, traffic generally increases in the following month. The listing of the MCHPE lecture "Studying Health and Health Care" on the Epidemiology Supercourse's 24 mirrored servers around the world will increase impact of the work, but complicate measurement of page hits [7]. Finally, we believe the internal pages are being heavily used by MCHPE staff; such usage is currently being measured.

Several national and international projects have relied on the knowledge repository fairly extensively. A large cross-national project based at Stanford University has used an index described in the Concept Dictionary (the Charlson Comorbidity Index) to help control for comorbidity before cardiovascular surgery. In another project, researchers in a five-province study funded by the Canadian Population Health Initiative will be operationalizing Continuity of Care as described in the Concept Dictionary

XSL•FO
RenderX

Figure 3. Number of Requests for Research Index Pages

Figure 4. Number of Requests by Unique Hosts

Figure 4. Number of Requests by Unique Hosts
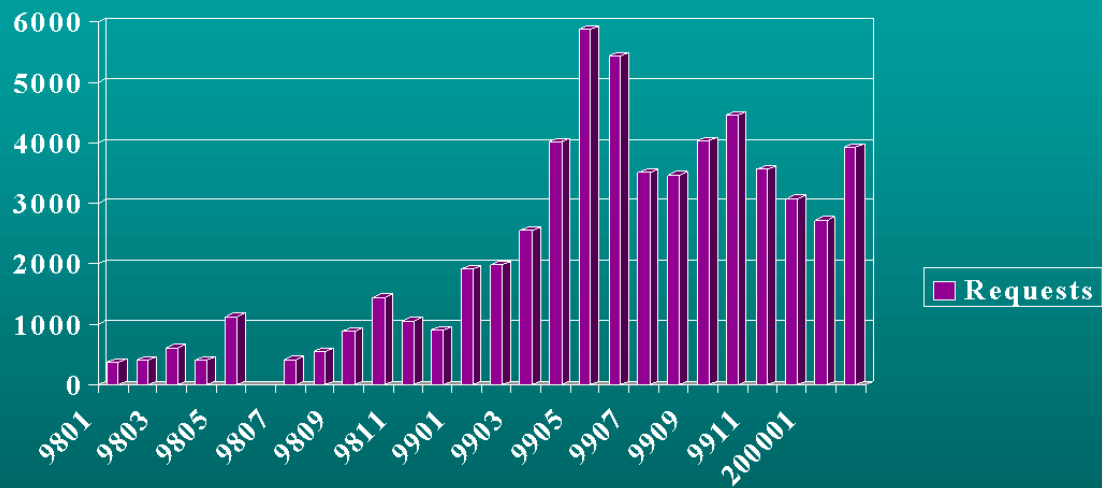
Year / Month Translation Example: 200001 represents January 2000
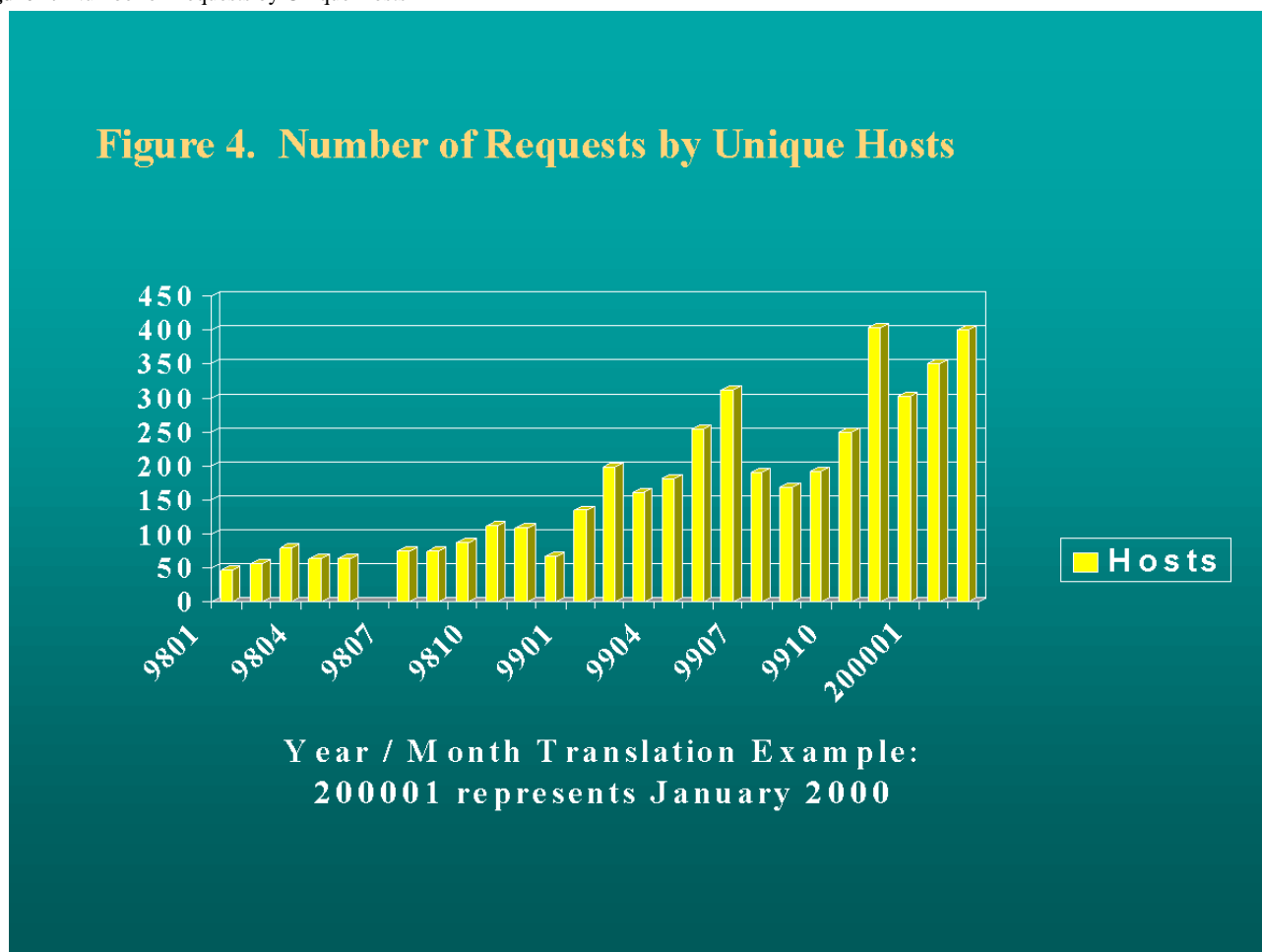
Discussion

Benefits of a Knowledge Repository

Organizing and structuring information into a knowledge repository can:

1. Promote standardization: An online information system encourages the use of standard terminology and methodology. The Concept Dictionary helps make explicit the assumptions associated with the development of a measure.

2. Increase efficiency: New projects at MCHPE often require complex programming and conceptual development. Our approach allows this intellectual property to be placed in the collective memory of MCHPE. Appropriate documentation, including descriptions of measures and detailed programming code, prevents duplication and protects the substantial time investment that such work involves. This process also keeps researchers current regarding concepts that may be modified from one project to the next.

3. Capture implicit knowledge: Capturing and making available tacit knowledge is a challenge for all organizations [5]. Frequently, a very small number of people know a great deal about the data or a specific concept. Such institutional knowledge needs to be saved in a convenient format to avoid missing important details and generating unnecessary effort to track down the expert on a specific topic. Making implicit knowledge explicit can lead to the discovery of a novel approach, or the identification of critical shortcomings in ongoing or completed projects.

4. Facilitate collaboration: Project-based websites, a recent addition to the online information system, can facilitate collaborations between MCHPE investigators and researchers from other centers. A website created for a three-province study, for example, lists the investigators from each of the provinces involved and provides a "What's New" section of project website updates. Overviews of each of the component studies include detailed methodology descriptions of data sources and eligibility criteria, as well as key concept definitions and analytic methods specific to the study. Relevant links to the Concept Dictionary and to external websites (such as the Canadian Medical Research Council and National Health Research and Development Program) provide users with additional project information.

5. Aid project management: Multi-site projects can obviously benefit from the use of web pages to coordinate the flow of information in a timely fashion. Operationalizing concepts using administrative data from different political jurisdictions is complicated by factors such as:

   a. Variations in data set characteristics and coding conventions. Not only do the contents of the data sets differ, the meaning underlying apparently similar variables may also differ among provincial data sets.

b. Varying demographics. An apparently straightforward decision in one setting may be complicated in another. For example, rapid suburban growth in Calgary, Alberta has made it necessary to interpolate between old census figures and projections from the latest results to define urban Calgary. In contrast, census data that are several years old present no special problems for estimating income quintiles in slow growth Manitoba.

c. Deal with conflicts between building on previous research and developing new approaches to resolve methodological problems: since the boundaries of Statistics Canada's enumeration areas do not always correspond with postal codes, appropriate census values for neighborhood income and educational statistics must be assigned to postal codes overlapping two or more enumeration areas. For the Manitoba postal codes, MCHPE researchers have assigned average income and education using population-weighted means from the overlapping enumeration areas. This procedure was utilized for 1986, 1991, and 1996 census data and designed to facilitate longitudinal research. Starting with 1996 census, Statistics Canada developed a different, probabilistic method of converting between enumeration areas and postal codes allowing a more consistent, repeatable method of assigning income and education values to individuals within these postal codes [8]. Results based on this methodology differed very slightly from those generated using population-weighted means, but nonetheless complicate cross-sectional and longitudinal comparisons.

6. Disseminate information: The MCHPE web-based system facilitates dissemination of current information to a wide range of users. The Glossary, in particular, is directed toward researchers who may vary significantly in their familiarity with technical terms. Links to other terms, combined with information about key "contacts," permit access to details not normally published in peer-reviewed articles. Income quintiles, for example, have been widely described in journal articles, but the specific methods for generating them have not been published [9]. The "Ongoing Projects" section permits users to determine (without waiting for publication) the current status of various projects undertaken by MCHPE researchers.

Future Directions

The knowledge repository may be characterized as a tool for increasing the flow of high quality information relevant to health services research. Meeting our objectives requires developing this resource in two directions: First, honing technical features will allow users who vary considerable in their expertise to more easily navigate through the available information. Second, the content must be continually updated and expanded to reflect the activities at MCHPE and related centers.

Considerable effort is being directed toward enhancing the number of links within dictionaries, from dictionaries to other sections of the MCHPE web site, and to non-MCHPE sites (click, for example, on PHRU Concept Dictionary; this was developed by the Population Health Research Unit at Dalhousie University). Extensive links among various glossary terms have recently been established (see, for example, A in the Glossary Index). We also hope to provide more direct access to concepts and definitions associated with individual papers by means of a search engine interface that returns all links (concepts, definitions and terms) used in a specific paper or report.

The content of the repository will expand as we incorporate new topics and new directions within MCHPE. Enhancing the linkages between geography and information (GIS) is an important direction for both research and dissemination. As MCHPE databases increase to include social service and educational information, terms and methods relevant to these new areas will be incorporated. We plan to include as much high-quality information as possible in our external web pages, while providing researchers with an informed sense of the issues involved in working across jurisdictions. Given our funding by Health Canada and various federal and provincial agencies, other researchers and organizations are permitted to copy parts of the dictionary. An advisory committee will generate suggestions for future additions, with considerable attention being given to what is most useful beyond Manitoba. Finally, we need to constantly address the tradeoff between adding content and organizing the information. Monitoring use of different entry points (Concept Dictionary, Research Definitions, Meta-Index, etc.) to the substantive material should help our investment decisions; considerable activity has been associated with each entry point.

Our efforts have been directed toward making documentation in the Research Index a routine component in project planning and development. The WWW, with its multi-dimensional capabilities for hot links and structured documentation, has provided a vehicle for meeting such objectives. As a knowledge base, the MCHPE online resource is new and possibly unique to health services research. We have only scratched the surface in terms of this resource's potential for training and education - locally, nationally, and internationally. The graduate course on the web site supported distance education in the fall of 1999. Visitors to MCHPE have already found the knowledge repository helpful in reviewing material before arrival.

Our experience is encouraging since accessibility to information, together with opportunities to receive feedback, moves us closer to an interactive format where research ideas can advance rapidly through an iterative process. With safeguards to assure information quality and increased opportunities for collaboration, we anticipate some real innovations in how health services research and epidemiology are conducted.

Acknowledgments

Conflicts of Interest

None

References

1. Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust N Z J Public Health 1999 Oct;23(5):453-459. [Medline: 20043017]
2. Roos LL, Romano PS, Fergusson P. Administrative databases. In: Armitage P, Colton T, editors. Encyclopedia of Biostatistics Volume 1. Chichester, UK: John Wiley & Sons, Ltd; Apr 1998:62-73.
3. Olson JS, Olson GM. Computer Supported Cooperative Work. In: Durso FT, Nickerson RS, Schvaneveldt RW, Dumais ST, Lindsay DS, Chi MTH, editors. Handbook of Applied Cognition. Sussex, UK: John Wiley & Sons; Jun 22, 1999:1-45.
4. Decoster C, Currie RJ, Turner D, Roos LL, Minish E. Communicating with the public, communicating with each other. Med Care 1999 Jun;37(6 Suppl):JS279-JS290. [Medline: 99314723]
5. Davenport TH, Prusak L. Working Knowledge: How Organizations Manage What They Know. Boston, MA: Project Management Institute; Dec 1, 1997.
6. ; SAS Institute. SAS Language. Cary(NC): SAS Institute; 1990.
7. Laporte RE, Akazawa S, Boostrom E, Campos M, Gamboa C, Gooch T, et al. Global public health and the information superhighway. Global health network university proposed. BMJ 1994 Sep 17;309(6956):737. [Medline: 95037920]
8. Wilkins R. PCCF+. Geocodes/PCCF Version 3 User's Guide. Automated Geographic Coding Based on the Statistics Canada Postal Code Conversion Files. Ottawa: Health Statistics Division, Statistics Canada; 1998.
9. Roos NP, Mustard CA. Variation in health and health care use by socioeconomic status in Winnipeg, Canada: does the system work well? Yes and no. Milbank Q 1997;75(1):89-111. [Medline: 97217280] [doi: 10.1111/1468-0009.00045]

Original Paper

Co-operative Health Information Networks in Europe: Experiences from Greece and Scotland

Efthimios Tambouris[1], MSc; M. Howard Williams[2], DSc; Constantin Makropoulos[1], Dr-Ing

[1]Division of Applied Technologies, National Center for Scientific Research "Demokritos", Athens, Greece

[2]Department of Computing and Electrical Engineering, Heriot-Watt University, Edinburgh, Scotland

Corresponding Author:
Efthimios Tambouris, MSc
Division of Applied Technologies
National Center for Scientific Research "Demokritos"
Agia Paraskevi, 153-10 Athens
Greece
Email: tambouris@nh.gr

Abstract

Background: Internet technology is transforming the general approach to communication and dissemination of information in the field of healthcare. However, it is also creating problems in terms of finding information, and knowing what credibility to place on the information found. The chaotic nature of the World Wide Web (WWW) and the simplistic approach adopted by search engines can make the task of finding relevant information difficult, and the user can waste considerable amounts of time on the process. Even when information is found, there is no general quality assurance process that can guarantee the credibility of the resulting information.

Objective: The aim of this research was to develop an approach for establishing co-operative health information networks (CHINs) with different focuses, which can be used in different parts of Europe. The resulting CHINs would provide organised healthcare information and support comprehensive and integrated sets of healthcare telematic services for a broad range of users. Such developments would reduce the difficulties of finding information and knowing what credibility to ascribe to it.

Methods: A common approach has been developed based on drawing together contributions from the major healthcare service providers in the region. Standard structures are recommended so that information is presented in a uniform way. Appropriate mechanisms ensure adequate security and a level of quality assurance for the end user.

Results: Since 1996, CHINs have been developed in six European countries as part of a European Union (EU) project. This paper presents the overall approach adopted, and the achievements in two different regions of Europe (Greece and Scotland). Although the circumstances in these two regions are very different, in both cases the resulting CHIN has proved successful.

Conclusions: CHINs offer a solution to the difficulty of finding relevant health information on the Internet and guaranteeing its credibility. They can be used in different ways in different regions, and have major benefits for both information providers and end users.

Introduction

Developments on the Internet have transformed our approach to communication and dissemination of information in a number of areas. Current approaches will develop further as the dramatic growth in Internet use continues. It is particularly relevant in the area of healthcare, where the ways in which information is published and accessed could change radically over the next five to ten years with significant consequences for both medical practitioners and patients. Co-operative health information networks (CHINs) represent one particular development of healthcare technology which holds considerable promise for the future. The CHIN project [1,2] started in Europe in 1996 with the aim of creating organised health information networks in different European countries; these networks were to be linked together to support comprehensive and integrated sets of healthcare telematic services for a broad range of users. The countries involved are Finland, Germany, Greece, Spain

(Catalonia), Sweden, and the United Kingdom (Scotland). Each region has different priorities; hence, a different slant to the CHINs has been developed in each case. This paper focuses on two of these regions and discusses the resulting developments.

In Greece, the focus has been on establishing a CHIN to act as a resource directory for health related information in Greece, and to provide health professionals with a number of telemedicine applications for remotely accessing multimedia patient records.

In Scotland, the focus has been on developing a publicly accessible CHIN with comprehensive coverage of a range of healthcare services. This is matched with a protected version that provides additional information which can only be accessed by healthcare professionals.

Methods

One of the overall aims of the CHIN project [1,2] was to develop a flexible approach to CHINs that could cater to a wide range of different requirements in different regions. In particular, an important distinction was drawn between online services that are intended primarily for health professionals, and those that are intended primarily for public access - by patients and the general public (as well as by health professionals on occasion). In the former case, an appropriate level of security is required, whereas in the latter security is not an issue.

Online services for professionals support various working scenarios between hospital staff and doctors in practices outside the hospitals, such as remote access to multimedia patient records, quality control for screening results, online patient referrals, and resource planning. All of these require a relatively high degree of security. In addition, there are online services to support access to a range of information: detailed information on local health services, professional education material, reference databases, statistical information and health data sets, etc. These require a lower level of security.

Online services for public access include web-based regional healthcare resource directories (a presentation platform for regional healthcare service providers), online consultations, patient education material, public health education material, information on support groups, and a wide range of information relating to healthcare services in the region. Technically, the approaches adopted are based on standardised, open, and scaleable solutions for computer and networking technologies, such as ISDN-based Intranets and HTTP (Hypertext Transfer Protocol); as well as for medical applications, such as DICOM and HL7. Users access the resource directories and the patient records via a standard World Wide Web (WWW) interface.

Approach in Greece

In Greece, the development of a CHIN was aimed at providing healthcare-related online services for public access and for healthcare professionals.

Online services for public access are provided through the development of a resource directory. Within the overall CHIN project, a standard approach has been agreed for such a directory. The Greek resource directory was set up in accordance with this approach, to provide a common platform for healthcare service providers to publish their content, and for Internet users to search for it. In the case of Greece, the penetration of Internet use is currently very small (only 1% in July 1999 by comparison with other EU countries, where the average was 20% in July 1999); nevertheless, it is expected to grow to comparable levels in the future. Not surprisingly, the amount of Greek healthcare-related content that is published on the Internet is sparse and often limited to a few pages. An exception is MEDNET (www.mednet.gr), a project by the Athens Medical Society; however, this resource concentrates mainly on information for the healthcare professional. For this reason, the resource directory in Greece was created to provide a focus and a motivation to healthcare service providers to publish their content on the Internet in an integrated fashion.

In its current form [3], the Greek resource directory consists of the following sections:

- Information on all hospitals in Greece,
- A detailed section on the National Health System (NHS),
- A pilot presentation on one disease,
- A section on medicine,
- A presentation covering telemedicine activities in Greece, and
- A section on the CHIN project.

All content (except for the NHS) is in both Greek and English, and the user can select the language of choice. The information presented on hospitals includes the contact details of each hospital in Greece. In accordance with the general approach developed within the CHIN project, the provided lists contain links to the web sites of all hospitals in Greece that have them (more than 15 in July 1999). In addition, each hospital and healthcare center that does not have a web site is provided with assistance to develop and publish its site on the CHIN server. A Java program named CHINBuilder, developed by the Scottish group, was translated into Greek and used to assist in generating these hospital web sites. All technical assistance and Internet costs are waived for public hospitals, which is one factor that has helped in achieving the widespread use of this resource. During the last year, four hospitals (Nikaia, KAT, Tzanneio, and Onassis Cardiochirurgical Center) used this service.

To the best of our knowledge, the detailed section on the Greek NHS is the only one available on the Internet and contains extensive material on Health, Welfare, and Public Insurance.

A pilot presentation on diabetes was developed to demonstrate and assess the capabilities of the WWW to provide health-related information to the Greek public. To this end, information on diabetes was compiled and presented. A particular focus of the presentation was to provide educational material for children. Included in this are a tutorial and a comic strip, which were translated into Greek and incorporated into the presentation with due acknowledgements. It should be noted that with the exception of some presentations in the pages of MEDNET, there is no similar information for patients available in Greek on the Internet.

The section on medicine provides a comprehensive list of related web sites in Greece. For example, links are provided to the

home pages of all medical departments of Greek universities. In this case, one of the initial objectives of the CHIN in acting as a resource directory for legacy information sources has been realised.

The section on telemedicine provides a reference point for all telemedicine activities in Greece. As in the case of the NHS, content had to be developed due to the lack of published material on the WWW about telemedicine in Greece. In addition to these services for public access, two online services for professionals have been developed. The first is an application that runs on a network connecting a hospital with a healthcare center, that allows the electronic exchange of medical results - especially for patients with diabetes [4]. These two health institutions are connected through leased lines to ensure the security of transferred data. This network will be connected to the National Diabetes Network at a later stage (see Figure 1).

Figure 1. The diabetes application
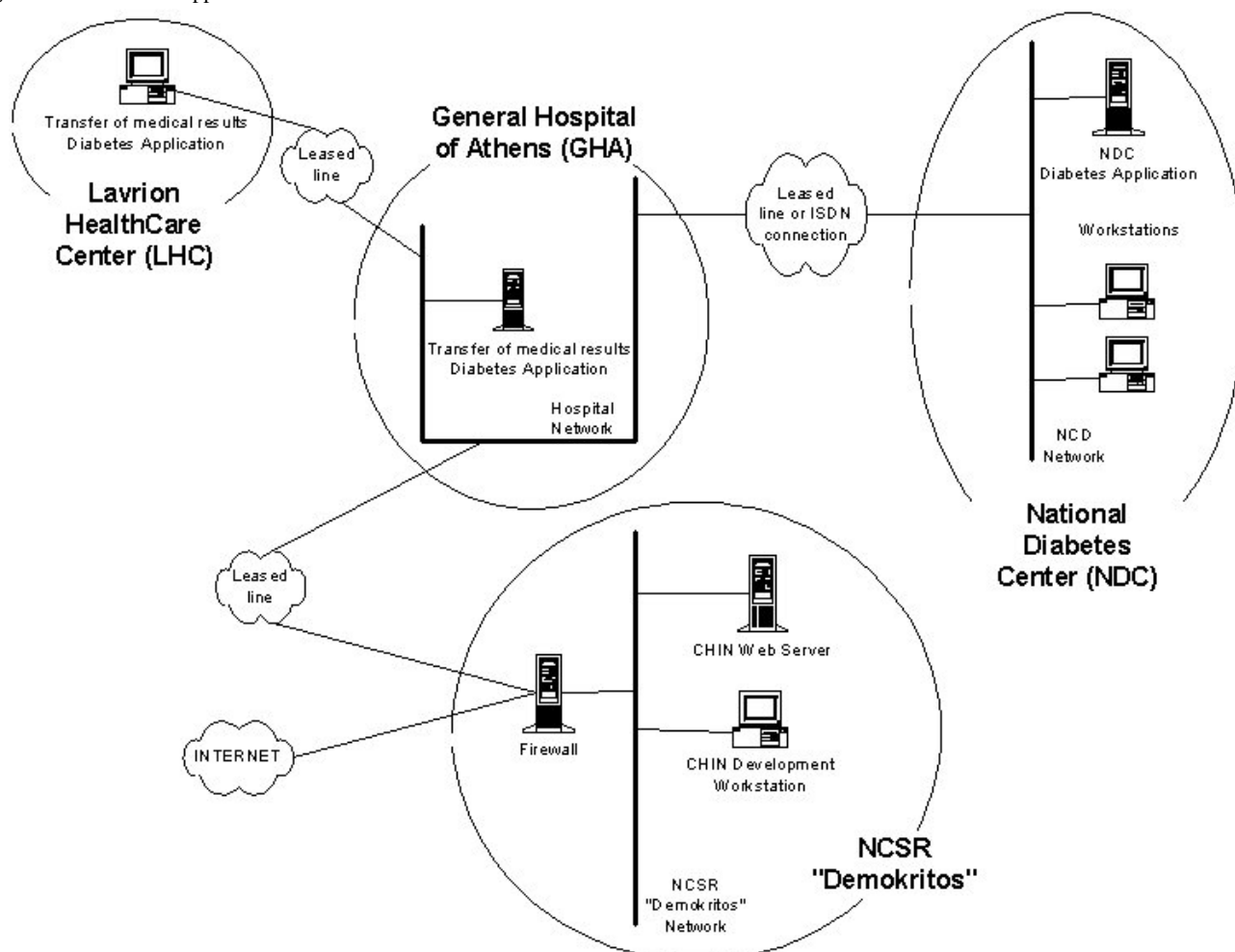
The second online service for professionals is a Picture Archive and Communication System (PACS) and a web-based application to access the stored images which has been installed in the General Hospital of Athens 'G. Genimatas.' The PACS installed was a product called DxMM and was provided by MedaSys. DxMM was installed at the Radiology department of the hospital. The system architecture is illustrated in Figure 2. Modalities are directly connected to Philips EasyVision, which also acts as a DICOM gateway. EasyVision is connected to DxMM and this is connected to WebMed [5], a web-based application that allows all workstations in the hospital to access the data. WebMed was provided by GMD (Gesellschaft fur Medizinische Datenverarbeitung mbH), another partner in the CHIN project. The hospital's network consists of fiber-optic cables between different buildings, and 100 MB/sec Ethernet within buildings. The system aims to create a filmless hospital.

Figure 2. Multimedia patient record application
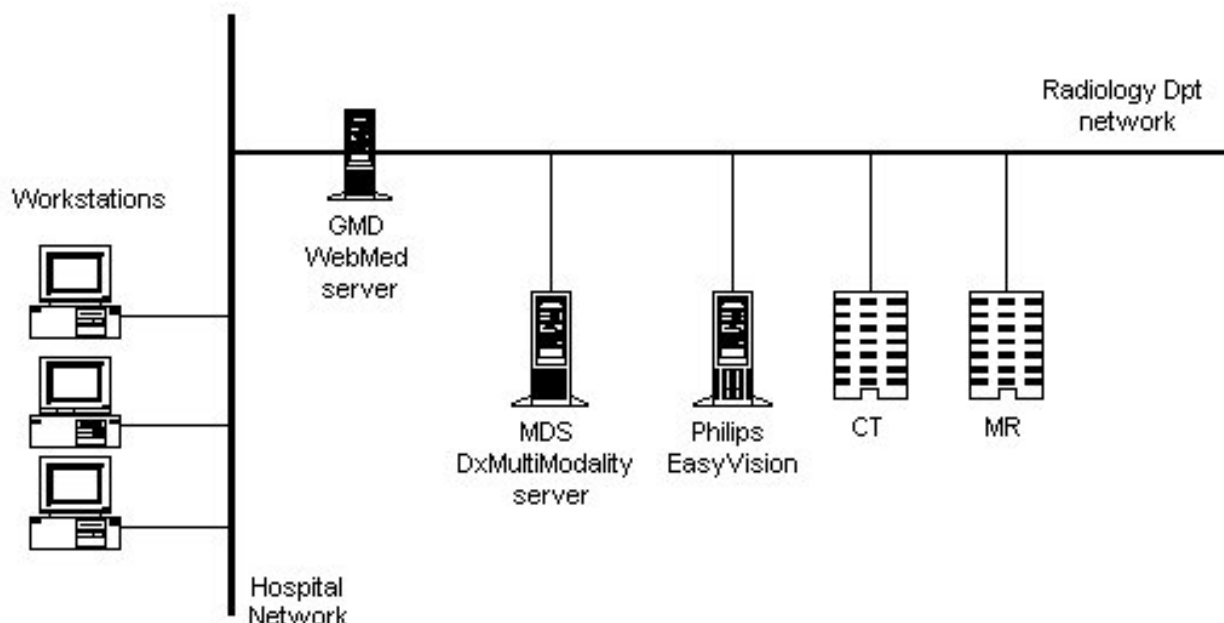
Approach in Scotland

The development of a CHIN for Scotland started with the idea of creating a publicly accessible CHIN that would bring together contributions from all major healthcare service providers in Scotland, to create an integrated collection of material aimed at patients and the general public. At the time, the thinking was a successful CHIN would draw in an increasing number of organisations, with a growing amount of material from each. As a result, it would rapidly become a local "reference library" on healthcare material, which was the obvious place for local citizens to refer to if they had an enquiry on healthcare matters. This would, in turn, increase the pressure on organisations to participate and to ensure that increasing amounts of information were published.

A major challenge of such a development is the organisation of the information structure [6]. It is essential that information is structured in such a way that as the CHIN expands and increasing amounts of information become available through it, the user will still be able to find what he or she is seeking with as little effort as possible. Needless to say, the user should be able to do so without getting lost in the potentially vast sea of information that could accumulate.

The first focus of development was on hospital trusts, and sought to establish agreement on the structuring of information provided by hospitals and on design guidelines that should be adopted in order to ensure consistency of presentation and a common look and feel. Although such technical aspects are the province of IT staff within the hospitals, commitment at management level was required in order to proceed.

In parallel with this, a second focus of development was undertaken in the area of health education. This covered both public health education and patient education material. The Health Education Board for Scotland (HEBS) has been responsible for providing the general public with information on topics of general interest (such as AIDS, drugs, cancer, heart disease, etc.) as well as advice on healthy lifestyles. Initially, the Board's developments focused on converting existing paper-based material to electronic form. For this they used a combination of static HTML (Hypertext Markup Language) pages, and dynamic pages generated from databases. However, HEBS rapidly moved on to multimedia presentations (including graphics, animations, and even video) designed for the WWW, which have greater appeal, especially with the younger generation. This has recently won them an award in the "Winners at the Web" competition in Scotland [7].

As the HEBS contribution has become established, other organisations have begun to provide information for patients on specific diseases (including particular types of cancer, asthma, diabetes, etc.).

A third focus of development was the area of general practice. However, here there was much greater reluctance to participate. Initially, offers were made to general practitioners (GPs) both through local GP committees and through a GP newsletter to create individual sites for free; but this produced no response at all. Two successive versions of a web site generator were produced to enable GPs to create their own sites with a minimum of effort. These were based on a set of templates, and a wizard application was used to lead the user through a sequence of steps to assemble and customise the templates to meet the user's requirements. However, this too met with little response. Finally, it has been agreed to dynamically generate web entries for the GPs by using information from the database of the National Health Service in Scotland; these web pages will contain the basic administrative details of all general practices in Scotland.

In addition to these three main focuses, there have been a large number of other health service providers that have added contributions, including:

a. Professional information - this includes information on Y2K compliance and other reference material, as well as direct links to databases such as the Travax database (which

b. provides up-to-date information on immunisations required for travel to every part of the world).

b. Professional education - this includes a set of medical guidelines for professionals provided by Scottish Intercollegiate Guidelines Network (SIGN), laboratory handbooks for professionals, etc.

c. Statistical data - the Information and Statistics Division (ISD) of the Common Services Agency has started to make its statistical data available through the CHIN.

The CHIN that has resulted from these developments is known as Scottish Health On the Web (SHOW) [8,9]. One of the objectives of SHOW is to maximise the benefits of individual contributors through integration. One aspect of this is to be able to move from one health service provider's contribution to that of another through the natural links in the system. This is clearly achievable through appropriate information structuring and indexing. A second aspect is encouraging contributors to make good use of cross-links between sites for the benefit of the user. A typical example is linking between hospital or general practice sites, and supporting public/patient education material. This is a longer-term development.

Another important issue is that of quality assurance [10]. Since it is clearly impossible for any central organisation to maintain proper quality assurance of all information provided by all contributors and ensure that it remains up to date, the responsibility is left with the provider organisations to establish proper internal controls over the content. However, in order to ensure that this responsibility is taken seriously, the information providers are required to display their logo on each HTML page so that the user knows who is responsible for the information. This is achieved through a standard frame-based layout, and information providers are offered support in producing layout. The idea of including the logo of the providing organisation on each page acts as an incentive to the management of the organisation to ensure that proper quality assurance controls are in place. At the same time, it provides an assurance to the user by clearly revealing the source of any information.

Most of the contributions are held on a small number of computers. There is tight control over these computers in order to maintain reasonable levels of security.

As SHOW grew, support for it was sought at higher levels, and the Management Executive of the National Health Service in Scotland (NHSiS) gave its backing. This has resulted in SHOW's acceptance as part of the NHSiS strategy for the future, and it has been mentioned in two government White Papers [11]. One consequence of this has been the development of two separate mirror sites, one public version accessible across the Internet, and one version with more available information, which can only be accessed by healthcare professionals via a protected network, NHSNet (see Figure 3).

Figure 3. Public and private components of Scottish Health On the Web (SHOW)
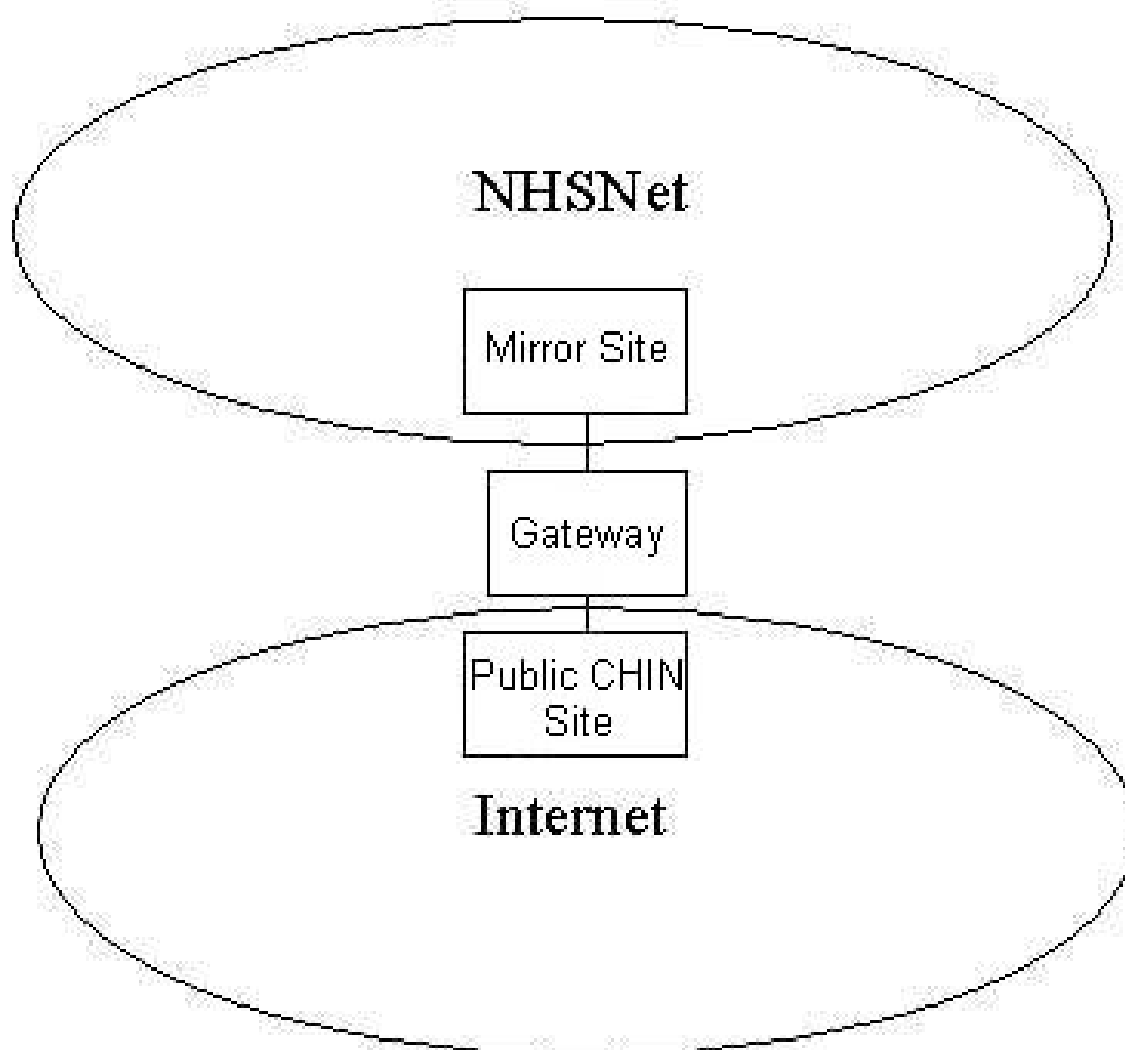
Results

The Greek CHIN server is one of the largest health-related web sites in Greece. The ultimate goal is to establish this site as the entry point for Internet users looking for health-related information in Greece. For this purpose, a number of activities have been initiated. For example, collaborations are being pursued with other healthcare related sites, such as the one developed by the Ministry of Health and Welfare. Web sites are being developed and included in the CHIN server free of charge for any Greek hospital and healthcare center that wishes to participate in this effort.

The Greek CHIN server has been recommended by a number of search engines and indexes in Greece. It is also the recommended health site by OTEnet, one of the biggest Internet Service Providers (ISPs) in Greece. The number of hits is growing, and in June 1999 there were almost 11,000 page hits (i.e. instances when a page was downloaded). An online evaluation questionnaire, developed by the CHIN consortium, is being used to evaluate the Greek CHIN. This questionnaire includes sections on the site's layout and navigation, as well as the content's comprehensiveness and usefulness; it further asks users for comments. Although at this stage there are insufficient responses for a comprehensive evaluation, the feedback thus far has been very positive. The main request from users is for more information on specific diseases and emergency procedures. The acceptance of the CHIN is also evident from the large number of incoming e-mails from Greek hospitals asking for guidance in publishing information on the WWW, as well as from professionals or medical students who are interested in the Greek healthcare system. Up to now, CHIN has assisted graduate students from Miami University (USA); a TV film producer (Sweden); postgraduate students at Exeter university (UK); and two EU projects, namely multimedia health information for citizens (MELIC) and tele-healthcare European network (THEN); among others. In summary, the Greek CHIN is trying to establish a service that is currently not offered by any other site in Greece, namely providing healthcare online information to the general public.

In terms of the online professional services, the initial goal was to familiarise health professionals with the new technology, with a view to establishing an integrated information system in the longer term. The long-term objective is to establish a slideless hospital, in which different information sources and databases containing images of patients from various modalities along with diagnostic reports, would be integrated and made accessible to all departments within the hospital.

The Scottish CHIN, SHOW, has as its overall goal the creation of a virtual healthcare library for Scotland, which will become the region's primary reference site for healthcare information. This will enable patients to take a more active role in their own healthcare, and provide support for professionals in a variety of ways.

Currently, of the fifteen health boards in Scotland, all but the four smallest ones have developed, or are in the process of developing, sites for SHOW. Of the 47 hospital trusts in Scotland, more than half have fully operational sites integrated into SHOW, or are preparing them [12]. All of the Headquarter organisations of the National Health Service in Scotland have SHOW sites that are fully operational or in preparation. In all, over 100 organisations now have fully operational SHOW sites. Yet others are still in a state of preparation.

The rate of access has grown steadily since the system first became operational and currently the hit rate is around 2 million page hits per month. Of this, over 90% of the page requests come from within the UK - unfortunately, it is not feasible to separate the statistics for Scotland from those of the rest of the UK, and thus we can take this no further at this stage. Assuming that this growth continues, the overall goal should soon be achieved.

As part of the CHIN project, CHINs are in varying stages of development in six regions within Europe. They can be accessed at the following Web addresses:

- Scotland (http://www.show.scot.nhs.uk)
- Greece (http://www.nh.gr/CHIN/)
- Catalonia, Spain (http://www.chc.scs.es/chin/inicial.htm)
- Brandenburg, Germany (http://b5www.berkom.de/chin/)
- Umea/Norrland, Sweden (http://www.cs.umu.se/~chin/index.html)
- Joensuu/North Karelia, Finland (http://www.pkshp.fi/english/eindex.htm)

The concept of Health Information Networks for Communities has been also addressed by a number of projects in the United States [13] and in Europe [14].

Discussion

In this paper, the concept of Co-operative Health Information Networks (CHINs) in Europe was presented. CHINs provide a common platform for healthcare providers to publish their content and for the general public to access it. In addition, CHINs provide a common platform for health professionals to support telemedicine applications.

The experiences from the development of CHINs in two parts of Europe, namely Scotland and Greece, were presented. The Scottish CHIN is recognised by the National Health Service in Scotland as the means by which to integrate information from most of the healthcare service providers in the region in order to provide the best service to users. The Greek CHIN started by producing content that was previously unavailable, and is now trying to expand by, for example, developing free sites for all public hospitals and by introducing telemedicine applications.

Although developments in different regions have diverged as needed to satisfy each area's different needs and priorities, collaboration at a European level has been beneficial to establishing each of the CHINs. For example, in Scotland the penetration of Internet use is much higher than in Greece, and consequently there is more health-related content available online. This has allowed developers of the Scottish CHIN to concentrate on the integration of content; Greek developers subsequently used this knowledge. At the same time, the Greek site has developed content for the general public, e.g. for children with diabetes, which can be used by Scottish developers.

Conflicts of Interest

None declared.

References

1. Williams MH, Venters GG, Venters GA. Towards co-operative health information networks. Bratislava: Slovak Academy of Sciences; 1995 Presented at: Proceedings of the 7th International IMEKO Conference, MBB95; 1995 Sept 6-9; Stara Lesna, Slovak Republic.
2. Weser AJ, Kleinholz L, Soderstrom PO, Eklund P, De Moor G, Williams MH, et al. Co-operative Health Information Networks. In: Proc MIE. Netherlands: IOS Press; 1996.
3. Co-operative Health Information Network, Athens, Greece. URL: http://www.nh.gr/CHIN/ [accessed 2000 Feb 14]

XSL·FO
RenderX

4.   Tambouris E, Makropoulos C. Application of telemedicine technologies for the diabetes care of patients in greek rural areas. Adv Clin Path 1998 Apr;2(2):163-165. [Medline: 10358356]

5.   WebMed. URL: http://www.gmd-net.de/englisch/press/index_press.htm [accessed 2000 Feb 14]

6.   Norman F. Organizing medical networked information (OMNI). Med Inform (Lond) 1998 Mar;23(1):43-51. [Medline: 98282404]

7.   Winners at the Web. URL: http://www.wow.org.uk/past/past98.html [accessed 2000 Feb 14]

8.   Venters GA, Williams MH, Venters GG. The SHOW Approach to Internet delivery of healthcare information. In: Proceedings of Current Perspectives in Healthcare Computing. Surrey(UK): BJHC Ltd; 1998 Presented at: ; 1998 Mar 23-25; Harrogate, UK.

9.   Williams MH, Venters GA, Venters GG. A National Information Network for Healthcare. Zagreb: KoREMA; 1998 Presented at: Proceedings of Eighth International IMEKO Conference on Measurement in Clinical Medicine & 12th International Symposium on Biomedical Engineering; 1998 Sept 16-19; Dubrovnik, Croatia.

10.   Impicciatore P, Pandolfini C, Casella N, Bonati M. Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home. BMJ 1997 Jun 28;314(7098):1875-1879 [FREE Full text] [Medline: 97367429]

11.   ; Scottish Office. Designed to Care: Renewing the National Health Service in Scotland. Scottish Office White Paper 1997.

12.   Williams MH, Venters G, Venters G, Marwick DH. Developing a Health Care Information System for Scotland. In: Proc MIE 1999. Netherlands: IOS Press; 1999 Presented at: ; Ljubljana, Slovenia.

13.   Payton FC, Brennan PF. How a community health information network is really used. Communications of the ACM 1999;42(12):85-89. [doi: 10.1145/322796.322814]

14.   Oates J, Jensen HB, editors. Building Regional Health Care Networks in Europe. Netherlands: Ios Pr Inc; Jan 1, 2000.

XSL•FO

RenderX