Enhancing Inter-Operability and Data Sharing In Medical Information Systems

Dhamir N. Mannai
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115, U.S.A.
(617) 437-3252 mannai@northeastern.edu

Khaled Bugrara
DYNEX, Inc.
Quincy, MA 02169, U.S.A.
(617) 479-2100 dynex@mcimail.com

ABSTRACT

Clinical care generates an immense amount of patient data that has been archived and manipulated by computer-based information systems. Such computer-based medical record systems improved the accessibility of clinical information and made several studies of such information possible. Unfortunately, the care provider's task of retrieving, integrating, and interpreting only those portions of the patient's record that are relevant to a specific clinical problem is actually becoming increasingly difficult. This difficulty can be attributed primarily to the large variety of minimum data sets, the heterogeneous formats used to store the data, the heterogeneous data access methods and procedures, the varying granularity of access to data, the different rigid views of the data, and the lack of inter-operability among the information repositories of such data sets. Recognizing the aforementioned issues, we are engaged in a project to build a multi-database environment tailored for the inter-operability of medical information systems. The main building blocks of such a system are a multi-disciplinary minimum data set and a catalogue for the support of interoperability and customization functions. In this paper, we report on the design approach used and describe the general architecture of the system.

I. INTRODUCTION

Currently, there are thousands of medical databases and medical data repositories. Such databases hold information considered valuable to clinicians, researchers, and administrators. Researchers need the data to support and validate research, healthcare providers need the data to support and enhance the quality of their services, regulators need such data and its associated statistics for decision support, and administrators need the data for financial and management matters. Even though such data has been collected, and automated procedures for data manipulation have been provided, accessibility to such systems is very often still inadequate. This inadequacy is due to the lack of flexible user interfaces to such systems, lack of a common ground for data exchange among the systems, and lack of software tools that assist data users to integrate the information repositories.

Given that data is a very valuable, hard to produce, and in some cases irreproducible resource, it is highly desirable to share it among the members of the medical community and to facilitate the way it can be accessed, retrieved, analyzed, and combined with other data. To accomplish this goal, we are establishing a multi-disciplinary minimum data set that serves as the common platform for data exchange among the various medical databases, and then we use it as the basis for providing inter-operability among the various medical databases. We also provide a user-friendly, user-customizable interface to the resulting system. The software system design is based on Open Systems and Inter-Operability approaches to provide efficient and, equally important, flexible access to information. By defining and promoting a multi-disciplinary minimum data set, interoperability, open systems interconnections, and the integration of various information sources will not only be possible, but will also lay the ground for future systems with high degrees of dynamic data interchange and interconnectivity (as opposed to simple static integration). Several significant advances made in real-time, heterogeneous, multidatabase management technology will then be readily adapted to medical information systems

SIGMOD /5/93/Washington, DC,USA
© 1993 ACM 0-89791-592-5/93/0005/0495...$1.50

once medical data formats and access methods have been brought closer to a common standard.

The approach used in this work is focused on the establishment of the multi-disciplinary minimum data set and on its use in providing inter-operability. There are already several efforts in the medical communities of North America, Europe, and the Far East aimed at establishing a standard for such a data set.

II. ISSUES IN MEDICAL INFORMATION SYSTEMS

The issue of computer-based patient record formats is especially important for our work since it represents the basis for all medical information systems. Recently, the Institute of Medicine commissioned a committee to study ways to improve the Computer-based Record [3]. The committee's study lead to the establishment of the computer-based Patient Record Institute (CPRI). One of the main functions of CPRI is to reexamine the Patient Record from all perspectives (e.g. definition, role, etc.) in light of the outstanding advances in computer technology and make recommendations for future formats of patient records. In an earlier but related activity, the American Society for Testing and Material (ASTM) proposed an extensive data model as the fundamental content of the Computer-based Patient Record independent of any clinical settings. The basic content includes inpatient/outpatient care information, long term care, trauma care, among others [2]. The proposed guide is known as Standard E1384 and is being adopted by the American Health and Information Management Association (AHIMA). At this time, AHIMA is sponsoring a field study for feedback on the proposed standard. The E1384 Standard basic data set is intended as a pool from which health care providers and researchers can draw longitudinal (summary) data views. A clinical research database that incorporates the E-1384 Standard data set is being developed at the University of Washington [5]; the database also supports the Uniform Hospital Discharge Data Set (UHDDS).

An emerging draft standard as a supplemental effort of the ASTM Computer-based Patient Record Project, is currently being developed to provide a uniform and clinically-sound longitudinal view of data [6]. Frisse [7] gives an assessment of the health of the Computer-based Patient Record research. In [4], Bradbury argues for the need of Computer-based Patient Record that is uniform, complete, flexible, expansible, and most importantly independent of any computer system constraints. He calls for the use of specialized natural language lexicons, and proposes a clinically oriented structure of the Computer-based Record.

Another important aspect of medical information systems that contributes directly to the degree of inter-operability among databases is the 'minimum data set'. The terminology 'uniform data sets' and 'basic data sets' first came to usage when the issues of health care were addressed in [8]. In clinical sites, the concept of a minimum data set is used to describe a relatively small number of data elements which each healthcare institution would collect. The goal is to provide the foundation for common and thus inter-operable medical information systems. There are numerous research publications that discuss issues relating to minimum data sets. Hutchinson [1] discusses the problems of deciding on a single minimum data set in Ambulatory Care. The Uniform Clinical Data Set System, which is sponsored by the Health Care Financing administration, uses a similar concept to study the quality of care in a uniform way throughout the US [9]. These data items, which include mortality, morbidity, readmission, and physician orders, are used to examine the quality of care provided.

Choudhury and Shah [10] describe a research oriented clinical information system in a tertiary care facility. The system is used to store basic data elements only (primarily due to constraints on disk space in the system.) The systems main function is to collect and use clinical data for prospective applications in clinical trails and individual therapy. In [11], Zwetsloo describes a four-step procedure on using hospital information systems for clinical research. This paper gives a good account of the information systems requirements in support of the clinical research. The study is especially interesting to us since its was carried out and documented by both clinical researchers and computer science professionals.

III. SYSTEM OVERVIEW

Today's medical computing environments consist of distributed, heterogeneous, and autonomous hardware and software systems. The distribution of these systems reflects the diverse nature of modern medical computing, while heterogeneity and autonomy arise in the process of fulfilling diverse information processing requirements. Originally, these systems ran in isolation to support their individual applications. Recently, it became evident that more complex applications involving multiple systems could be supported only if the systems could cooperate with each other to perform a multi-system task. Unfortunately, the potential for cooperative interaction was not considered in their original design and there is as yet no general model supporting inter-operability among these isolated systems. As a result, the processing of global applications (such as studying disease trends in the community) that involve multiple systems becomes a formidable task and remains largely manual.

Multidatabase applications require services from multiple

systems. To support such multi-database applications, our software system provides the following capabilities:

(1) Access to multiple heterogeneous databases through a unified interface. This capability will free users from involvement with system-specific details.

(2) Ability to specify multi-database applications and the flow of their execution. Through the multi-database software system, users will be able to specify both data and control flow among the different systems accessed for the purposes of executing the application.

(3) Ability to 'inquire' about the structures of participating databases, their data contents, and the quality of their data. To support this capability, a catalogue of databases is needed.

(4) Freedom of association and controlled access. Participating organizations shall be able to join the multi-database federation (by putting their database on-line for access) and leave it as they see fit and at any time. Furthermore, they must be able to control access to their information.

The Multi-Database System: The system being developed consists of a collection of databases accessible through a common graphical user interface. Various users see different views (external schema) of the stored data, and have different customized interfaces to the system. The system provides the mechanisms necessary to access information remotely, extract information from different databases, and support multi-database applications in accordance with the objectives outlined above. In addition, we provide a user-friendly, user-customizable, graphical user interface through which the user interacts with the multiple databases. The overall architecture of the system is depicted in Figure 1.

The center piece of the system is the graphical user interface that hosts the data import tools, the multi-database applications generator, statistical packages' interfaces, and the execution controller of multi-database application programs. Through the user interface, users can browse through descriptions of component databases and their data attributes, generate application programs for multi-database access, and invoke statistical operations on retrieved data. The Multi-database Access Program Generator (MAPG) is an interactive tool that assists users in writing programs that access multiple heterogeneous databases. Users can simply indicate their informational needs and the databases to be accessed with no concern for the actual access methods. Access methods and location transparency capabilities are automatically resolved by the software system. Application

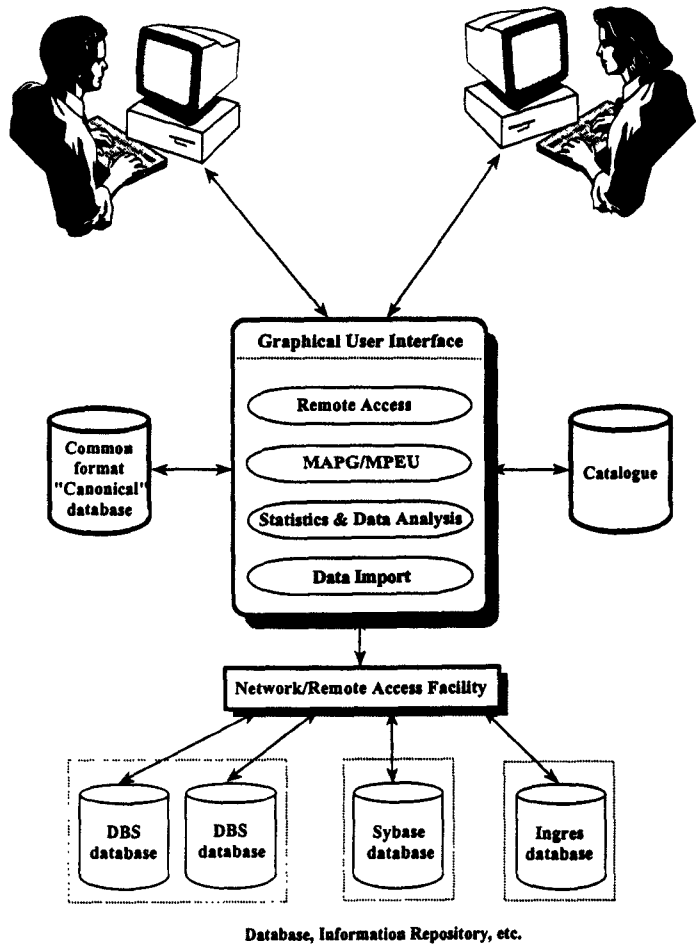

FIGURE 1 GENERAL SYSTEM OVERVIEW

programs generated using the MAPG can then be submitted to the Multi-database Program Execution Unit (MPEU) for processing. The MPEU coordinates the tasks needed to execute the application and present its outcome to the user. Implementation details of the component databases, their location information, and their structures are stored in a catalogue. The multi-database software system relies on the catalogue to provide users with a unified interface.

Computer Catalog of Medical Data Repositories: There are currently hundreds of databases that contain medical information. Such databases are geographically disperse, controlled by different organizations, hosted on heterogeneous systems, managed by different software systems, and reside on different media. Furthermore, such databases may support different data sets, common data sets with different formats, common information with varying degrees of data accuracy, data reliability, and data completeness, etc. Software tools that access such systems and extract information from them need to know a number of parameters about them. In order to provide an automated systematic approach to obtain the information

about such databases and their characteristics, a 'database of databases' is needed. We call such a database the Computer Catalog of Medical Data Repositories. As part of this project, we are building such a catalogue of databases. For each data repository, the Catalog contains descriptions of the participating databases in terms of their contents, the quality of their contents, and the accessibility to the data. The database attributes maintained by the Catalog include name of the database, agency in charge of its administration, location (internet address) of the database, the information it contains, and ways to access it. For each data collected in that database, the Catalog contains indications as to its reliability, accuracy, completeness, and any other quality measure used by the medical community. In addition, it includes administrative attributes such as the cost of obtaining the information, authorizations and security clearances needed, and possibly some assessment of the overall state of the database. One of the main objectives of the catalogue is to automate the data import functions and to assist in the formulation and processing of multi-database application programs. The catalogue includes mappings of the attributes of the participating databases to user-defined functional data sets. This mapping may be stored either in the catalogue or in the user's profile description.

Multi-Database Query Processing: A multi-database query is, in fact, an application program that requires access to multiple databases as part of its execution. Our system offers the capability to write application programs that span multiple databases potentially residing at multiple sites. The user need not be concerned with the location of databases and the ways to access them. Once an application program is generated (using the Multi-database Access Program Generator), it is submitted to the Multi-database Program Execution Unit (MPEU) to coordinate its execution. The MPEU consults the Catalogue for location information and access methods and then decomposes the application program into units. Each unit will be destined to one participating site. At the participating sites, there will be Remote System Interface modules (daemons) ready to process the incoming program units. Upon execution of the various units, the RSI modules will transmit their results to the MPEU at the user's site. The RSI modules are now being phased out in favor of providing simple terminal connection sessions to participating databases.

IV. CONCLUSION

The most difficult problem in this project is expected to be the lack of consensus on what a medical minimum data set is. However, whatever the outcome, there will be a great deal of common-platform agreement among the various members of the health profession that enables us to provide significant interoperability. Although this work is limited to clinical databases, the techniques should, in principal, generalize to other branches of medical computing in particular, and to the physical sciences in general.

REFERENCES

[1] Hutchinson, A., "Designing European Minimum Data Sets for Ambulatory Care", Med. Info. Vol 16 No. 1, 1991.

[2] Medical Computing Committee E31, "A Review of Published Standards and Current Projects", Feb (1992), ASTM.

[3] Institute of Medicine, "The Computer-Based Patient Record: An essential technology for health care", Washington, DC, National Academy Press (1991).

[4] Bradbury, A., "Computerized Medical Records: Defining a Standard Without The Computer", Med. Info. Vol 16 No 3, 1991.

[5] Hanken, Mary Alice and Forrey Arden, "Construction of a Common Clinical Data Catalogue", Journal of AHIMA Vol 63, No 8, Aug 1992.

[6] Gretchen, M., "Patient Information Standards Development: A Progress Report from ASTM", Journal of AHIMA, Vol 63 No. 8, Aug 1992.

[7] Frisse, Mark, "The Health of The Computer-based Patient Record", Medical Academia, Vol 67 No. 7, July 92.

[8] Muranghan, J., H., "Long Term Care Data", Medical Care, Vol 14 No. 5, Supplement 1976.

[9] The Health Standards Quality Bureau, "The Uniform Clinical Data Set", Journal of AMRA, Vol 62 No. 7, July 1991.

[10] Choudhury, M. and Shah, M., "PC-Based Research Oriented Clinical Information System", Med. Info. Vol 16 No. 4, 1991.

[11] Zwetsloo-Schonk, J., et al., "How to Approach a hospital Information System as Sampling Frame", Med. Info. Vol 16 No 3, 1991.

[12] McDonald, C., Blevins, L., Tierney, W., and Martin, D., "The Regebstrief Medical Records", MD Comput. Vol 5 No 34, 1988.

[13] Barnett, G. et al., "COSTAR System", Information Systems for Patient Care (Blum, B. editor). Springer-Verlag, New York, 1984.

[14] Coira, E., "Incorporating User and Dialog Model into the Interface Design of an Intelligent Patient Monitor", Med. Info. Vol 16 No 4, 1991.

[15] Sheth, A., Larson, J., "Federated Distributed Database Systems for Managing Distributed, Heterogeneous, and Autonomous Systems". ACM Computing Surveys, Vol 22 No 3 (1990).

[16] Litwin, W., Leo, M., and Roussopoulos, N., " Interoperability of Multiple Autonomous Databases". ACM Computing Surveys, Vol 22 No 3, (1990).

[17] Elmagarmid, A.K. et.al, "Transaction Specification in Multidatabase Systems Based on Parallel Logic Programming," International Workshop on Interoperability in Multidatabase Systems, Kyoto, Japan, 1991.

■■