The Columbia Medical Informatics Story: From Clinical System to Major Department

James J. Cimino, M.D.

When Paul Clayton was recruited in 1987 to become director of the Center for Medical Computing at Columbia University, his vision was to do more than just buy and install clinical applications software. Columbia was one of the recipients of the original IAIMS (Integrated Academic Information Management System) grants from the National Library of Medicine, and as such was seeking to develop a proposal for a five-year implementation grant. At the same time, the Presbyterian Hospital—Columbia's principal affiliate hospital and neighbor on the Columbia-Presbyterian Medical Center (CPMC) campus—wished to pursue the development of its own clinical information systems. The hospital president, Dr. Thomas Morris, and the medical school dean, Dr. Henrick Bendixen, found a natural synergy between these two goals and thought of Clayton from the University of Utah's Department of Medical Informatics (DMI), at the time, the only such department in the country.

At Utah, Clayton had been one of the principal architects of the Latter Day Saints Hospital HELP system (an advanced hospital information system that includes automated decision support) and as a condition of his recruitment, he insisted that the University establish a true academic department—one that would conduct research and develop the systems that CPMC needed. Clayton's vision was to combine the support of the hospital, university, research grants (including IAIMS), and industry partners to build the next generation HELP system. He saw this system as one that would serve as an enterprise-wide repository to support a variety of clinical needs, including automated decision support. So he gathered several key information systems personnel from Columbia and recruited several others (including me). What started as an effort to build a clinical system has grown into one of the world's largest research and teaching programs in medical informatics. In 1994, the university promoted the center to departmental status.

Service as the Centerpiece

The service responsibilities of an academic department are often the least publicized aspects of its work. At Columbia, however, it is the centerpiece. Many components of the system were developed as part of various research projects, and the system as a whole serves as the basis for many others discussed below.

In designing Presbyterian Hospital's clinical information system, we chose to develop a central repository for patient data in which we would store information from whatever sources were available, and then provide ways to retrieve and view it. The basic architecture consists of five main components:

What started as an effort to build a clinical system has grown into one of the world's largest programs in medical informatics.

Central Repository. This is a large relational database that captures all data coming from ancillary systems. Data can be text reports, structured coded data, or images. It is patient-oriented (as opposed to encounter-oriented) so that it can serve as a longitudinal record for patients regardless of whether they are inpatients or outpatients. The relational tables are designed in a row-oriented, rather than a column-oriented or table-oriented manner; that is, data are stored with a code for the type of data in one column and the actual data themselves in a second column. This allows us to add new data types (such as occurs when a new ancillary system starts sending data to the repository) simply by creating new codes for them, rather than having to create new tables or add columns to existing tables.

Data Access Modules (DAMs). These are programs that allow external applications to store and retrieve data from the central repository. Each performs a specific query or update, such as storage of laboratory results or retrieval of a recent radiology report. Queries can be made for coded data (such as "Get me all the whole blood glucose test results"), and they can be based on a class of codes (such as "Get me the medication orders where the drug is an antibiotic").

Medical Entities Dictionary (MED). This is a repository for all the coding terminologies used by senders of data. Terminology servers on several different platforms provide information about these terminologies so that, for example, data being uploaded to the repository can be translated by the upload program prior to storage by the DAMs. The MED is also used to convert coded data back to textual form (for example, for displaying laboratory results) and supports a variety of queries for terminology information, including classification information (as in "What drugs are antibiotics?").

Data Monitor. This program is responsible for storing data provided by the DAMs and for checking the type of data to see if any special logic should be evaluated. The logic is in the form of Medical Logic Modules (MLMs) written in the Arden Syntax, a standard for representing medical logic. An MLM defines conditions by which the Data Monitor should trigger its evaluation, after which it can query the repository for additional information needed for the logic. For example, an MLM might set the trigger as "storage of a serum creatinine value"; and when such an event occurs, the MLM might check the current value and then query for previous values to determine if a potentially dangerous trend in results (signifying impending renal failure) might be occurring. If the logic detects the condition, it generates a warning message that is stored in the repository, and also issues warnings by E-mail, fax, or phone.

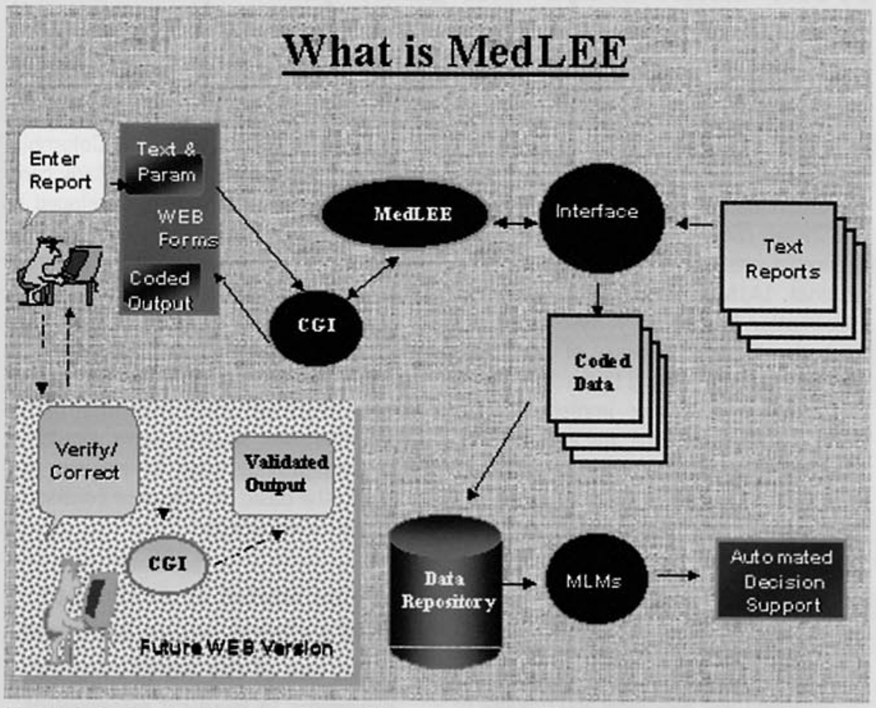
Medical Language Extraction and Coding System (MedLEE). MedLEE is a natural language-processing system that takes text reports being stored in the repository and generates structured, coded data that can be stored in the repository. These data are particularly useful for the decision-support component of the Data Monitor (see figure).

In addition to data storage, we provide a variety of users with access to the data for a variety of purposes. For many years, the principal method for access was an application simply called CIS (for "Clinical Information System") which contained a set of display screens for looking at a patient's data. For a variety of reasons (including accessibility and Y2K problems with its operating system), CIS is being phased out in favor of a version created for use on the World Wide Web called WebCIS.

Other systems have been created to support institutional functions, from storing a doctor's problem list to printing prescriptions to supporting the house staff's nightly cross-coverage. Each of these makes use of the DAMs via the Internet to carry out the desired operations. For example, a system for use in the clinics called DOP (for "Doctor-supported Outpatient Practice") displays data from the repository and accepts new data such as medications, problems, allergies, vital signs, and visit notes. Most of these systems will be folded into WebCIS.

The service responsibilities of the department have recently been expanded to accommodate the changes associated with the merger of Presbyterian Hospital with New York Hospital, which formed the New York Presbyterian Hospital (NYPH). While each campus will continue to have distinct information needs in the near term, there is an overarching need to have a single enterprise-wide clinical repository so that doctors from one institution can obtain relevant information about their patients, regardless of where the information was collected. Recent affiliations between NYPH and other healthcare organizations have included provisions for expanding this repository to encompass information collected at some 37 institutions.

Fortunately, our architecture has been readily adaptable to this expansion. Although some basic changes were necessary to accommodate institution-dependent patient identifiers and allow each institution to segregate certain data for confidentiality reasons, the basic components of the system have remained the same. For example, the laboratory terms from the New York Hospital laboratory system were added to the MED in the appropriate classes such that a query for "Get me all the blood sugars" will retrieve all pertinent results, regardless of the originating laboratory.



MedLEE, developed at Columbia, is one of the components of DMI's central repository for patient data and is breaking new ground in language understanding. A natural language processing system, MedLEE allows patient reports to be extracted, structured, and encoded so the data can be used in applications such as automated decision support. Illustration by Carol Friedman, Department of Medical Informatics, Columbia University.

Research as a Critical Link

The DMI currently comprises nine full-time faculty, seven adjunct faculty, and 47 fellows and students. There are many inextricable links between our research and service work. In some cases, the research has been used in practical settings to serve as key components of the hospital's information architecture. In other cases, the clinical system has provided invaluable resources for supporting research projects: Users have made themselves available as test subjects, and the hospital has been a living laboratory. In still other cases, the onerous tasks associated with maintaining the clinical information system have led to the development of better solutions to common problems. These links are many and varied, as are the types of research we carry out. The following are descriptions of some of our projects:

Database Modeling. The design of the NYPH clinical repository was a marked break from the type of repository found in financially oriented healthcare systems. The "row-oriented" approach yielded a database that has proven to be easier to expand and allow more versatile queries. Our design led to our winning the Computer-Based Patient Record Institute's first Davies Award. Since then, research has focused on developing user interfaces for constructing queries that take full advantage of the database design rather than research on the database itself.

Decision Support. The Decision Support system, embodied in the Data Monitor, was created using the Arden Syntax for medical-logic representation. The development of the syntax was, itself, a collaborative research project involving many groups from many countries to develop a language for sharing decision-support logic. The Arden Syntax came together at a Columbia-hosted conference held at the Arden Homestead Conference (hence, the name), in Tuxedo, New York. Since then, the research has expanded to bring decision support to more and more venues. For example, a project is currently under way to bring decision logic to bear on the problem of immunization recommendations. This project also exploits the clinical repository as a convenient place to store all patient data. Not only is this convenient for the immunization project but also for those who require all sorts of patient data for many (approved) uses.

Controlled Terminology. The MED is more than a simple collection of coding terminologies. In addition to the standard information needed to allow the proper use of a term (such as its name and code), the MED contains conceptual knowledge. This has enhanced our ability to translate terms from one terminology to another. It also supports sophisticated "inferencing" for use in decision support and terminology maintenance. For example, the knowledge in the MED has been used to merge terms from multiple laboratory systems into a single, coherent classification scheme and has even been used to propose new classes for grouping similar test terms together. The ability to traverse the MED's semantic network has been used to support the inferences needed for "infobuttons" (see below).

Natural Language Processing. Although used as a way to derive coded information from text reports in the repository, MedLEE is a cutting-edge natural language processing system that is breaking new ground in language understanding. For example, in one recent study, MedLEE was shown to be as good as expert physicians at identifying significant findings in radiology reports.

Distributed Systems and Security. The complex nature of medical practice settings, coupled with the severe restrictions necessary for patient privacy and confidentiality, result in extremely difficult authorization problems. Who is allowed to look at which patient data? The answer is not always easy, and is always changing—influenced both by shifting legal requirements and by changing practical considerations, such as physician cross-coverage. We are developing sophisticated models for accommodating the need for privacy and legitimate needs for access. In addition, we are building intelligent monitoring systems that will raise a red flag when the activity associated with a particular user changes significantly from the individual's typical behavior patterns (for example, suddenly looking at large numbers of patients or at patients from different specialty services). In other words, it can detect an imposter who is gaining access with a stolen password.

Data Mining. The clinical repository is designed to support patient-oriented queries (for example, "Get me this patient's blood sugars"). We have also developed a second database that is optimized for cross-patient queries (for example, "Find me all the patients with elevated blood sugars"). This database supports the traditional needs of clinical research, but is also being used for data mining to find new patterns and correlations among clinical data sets.

Telemedicine. The ability to perform healthcare tasks from a distance is becoming more feasible with the advent of new technology for information collection (including inexpensive video cameras), combined with the ubiquity of the Internet. We have several projects underway to explore the idea of merging these two technologies—for example, by using a hand-held device consisting of a spirometer and a computer, asthmatic patients can record their respiratory data at home and transfer the results by phone (even cellular phone) to our repository. Once there, technicians can view the results with a Web browser and the Data Monitor can alert physicians when worrisome measurements are evident.

We are building intelligent monitoring systems that will raise a red flag when a particular user's activity changes significantly. In other words, it can detect an imposter who is gaining access with a stolen password.

Infobuttons. Studies have shown that the information needs of clinicians are often unfulfilled. The Internet is a powerful medium for accessing information resources and has great potential for addressing this problem. We are developing ways to tie WebCIS to Internet-based resources through simple links called "infobuttons" (so named for the international "i" icon we use to mark the links in WebCIS). The infobuttons can make use of coded data in the repository and MedLEE to extract terms from text reports. Using the MED, these data can then be translated as necessary to serve as input for the various resources we have available. Some examples include: linking medication orders to the relevant pages of the Physician's Desk Reference, translating organism names from microbiology reports into Medical Subject Headings (MeSH) terms to search Medline, and extracting findings from x-ray reports to search a radiology library in England.

Patient Access to Health Information. Given our newfound ability to do almost anything on the Web—from shopping, to banking, to travel reservations, to tracking express packages—it is only a matter of time before reading our medical records online becomes commonplace. Yet, little is known about how patients use and interpret their records. We are exploring this with a simplified version of WebCIS (called PatCIS) that provides patients with a customized view of their own data from our repository. We are enhancing the displays by adding patient-oriented infobuttons and guidelines. By studying PatCIS users, we hope to learn how to improve patients' understanding of their own conditions and improve communication between them and their physicians.

Teaching as a Mission

The Columbia DMI faculty has always viewed teaching as part of its mission. We started with a few isolated courses and a weekly seminar, but formal programs gradually evolved. Although these have entailed greater teaching responsibilities, they have also provided us with a ready supply of assistants and collaborators for research projects and service work.

Masters of Arts (M.A.) Program. This program is intended to provide students with a broad-based understanding of the issues and aspects of the field of medical informatics, and some depth in such specific areas as decision support, data analysis, systems architecture, and user interfaces. A weekly seminar and journal club provide opportunities to learn about—and discuss—the latest research developments at Columbia and elsewhere. Students participate in one project and one major writing assignment (a Masters Essay), which is intended for publication to provide some experience that will prepare them for careers in industry or academia. The course work can be completed in a single year on a full-time basis, or in two or more years part time.

Doctor of Philosophy (Ph.D.) Program. This program trains students for more advanced careers in medical informatics. It fulfills the requirements of the M.A. degree and supplements it with additional course work, project work, teaching experience and, of course, a thesis. Students complete their course work in the first two to three years while participating in projects as a member of one or more of the

Given our newfound ability to obtain almost anything via the Web—it is only a matter of time before reading our medical records online becomes commonplace.

DMI's research groups. The transition to the "thesis phase" of the program is marked by satisfactory performance on all six parts of a comprehensive exam. The design and completion of the thesis is accomplished in one or more (typically two) subsequent years, with doctoral candidates given tuition and stipend support from training and research grants.

Postdoctoral Fellowship. The fellowship is intended to prepare those with an M.D. or Ph.D. for careers in medical informatics research. Although fellows perform some course work and attend the seminar and journal club, their primary focus is on research. This work is initially done as part of some larger existing project with fellows eventually pursuing interesting research questions of their own. In addition, each fellow is expected to take on some service obligation such as maintaining a specific subcomponent of the clinical information system. Most of the fellowship positions are funded by the National Library of Medicine, with additional positions funded by the National Institute of Dental Research, the New York State Psychiatric Institute, and a variety of individual funding mechanisms. Those funded by the federal government are restricted to U.S. citizens and permanent residents.

Striking a Balance

Columbia's Department of Medical Informatics has found a balance between service, research, and teaching in which each aspect of the triumvirate benefits from the others. There is a substantial measure of pride in the fact that this department was built from the ground up, with one idea evolving into a multifaceted educational entity. During that evolutionary process, we have found that the architecture we pursued over ten years ago has not locked us into the past. Instead, it served as the platform from which we could begin to address the new information challenges in healthcare. ●

James J. Cimino, M.D., is associate professor of medical informatics and medicine, Columbia University College of Physicians and Surgeons.

Roger Dahlen, Ph.D., Consultant, is section editor for this column.

Additional Reading

1. Hendrickson G, Anderson RK, Clayton PD, Cimino JJ, Hripcsak G, Johnson SB, McCormack M, Sengupta S, Shea S, Sideli R, Roderer N. The integrated academic information management system at Columbia-Presbyterian Medical Center. MD Comput 1992; 9(1): 35–42.

2. Clayton PD, Sideli RV, Sengupta S. Open architecture and integrated information at Columbia-Presbyterian Medical Center: The role of networks, workstations and databases. MD Comput 1992; 9(5): 297–303.

3. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Informatics Assoc 1994; 1(1): 35–50.

4. Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. Computers Biomed Res 1994; 27: 291–324.

5. Friedman C, Alderson P, Austin J, Cimino JJ, Johnson SB. A general natural language text processor for clinical radiology. J Am Med Informatics Assoc 1994; 1(2): 161–174.

6. Johnson SB, Forman B, Cimino JJ, Hripcsak G, Sengupta S, Sideli R, Clayton PD. A technological perspective on the computer-based patient record. In: Steen EB, ed. Proceedings of the First Annual Nicolas E. Davies CPR Recognition Symposium, 1995 April 4–6; Washington, DC. Washington, DC: Computer-Based Patient Record Institute, 1995: 35–51.

7. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Medicine 1995; 122: 681–688.

8. Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: application development using the World Wide Web. J Am Med Informatics Assoc 1995; 2(5): 273–284.

9. Cimino JJ, Allen BA, Clayton PD. Medical informatics training at Columbia University and the Columbia-Presbyterian Medical Center. In van Bemmel JH, McCray AT, eds. Yearbook of Medical Informatics. Rotterdam: International Medical Informatics Association, 1995; 125–129.

10. Johnson SB. Generic data modeling for clinical repositories. JAMIA, 1996; 3(5); 328–339.

11. Hripcsak G, Clayton PD, Jenders RA, Cimino JJ, Johnson SB. Design of a clinical event monitor. Comput Biomed Res 1996; 29(3): 194–221.

12. Cimino JJ, Elhanan G, Zeng Q. Supporting Infobuttons with terminological knowledge. J Am Med Informatics Assoc 1997; 4 (Suppl): 528–532.