Integration and Interoperability of a Multimedia Medical Distributed Database System*

Alfonso F. Cardenas‡ Ricky K. Taira† Wesley W. Chu‡ Claudine M. Breant†

‡ UCLA Computer Science Department
and
† Department of Radiological Sciences
University of California, Los Angeles

1 INTRODUCTION

Medical practice and research now rely heavily on the use of information stored in computer systems. Due to the many specialized branches of medicine, these database systems have developed independently, their design reflecting the innovativeness of the database implementors, the scope of data required for their operation, and the culture of their department. Each user group chooses its own database management system (DBMS) according to its needs. Example medical databases include: 1) hospital information systems (HIS), 2) radiology information systems (RIS), 3) picture archiving and communication systems (PACS), and 4) various research systems. These independent database systems provide only a first order solution to the management of the large repository of image, text, and scientific information generated by academic medical research centers. Research involving data stored in multiple databases often is hindered by the fact that these databases operate under different operating systems, have different data access and manipulation languages, have different data semantics, and have different communication protocols. Furthermore, researchers investigating similar areas of study often have no clue as to existence of databases which contain information that may accelerate or substantiate their own research. Figure 1 shows various medical data repositories present at our institution.

In addition to providing the researcher with more data, scientific medical databases require better query answering capabilities. Epidemiology and biological modeling studies (e.g., human genome mapping, functional brain mapping) are concerned with correlating image features (radiographic, pathology, etc.), patient symptoms, and diagnostic performance measures with underlying disease states. These correlations will vary with respect to patient profile (sex, race, diet, genetic composition, etc.). Furthermore, these mappings change with time and may evolve with respect to patient maturation processes, disease progression, and therapeutic events. This type of medical research requires the ability to access images by content rather than by artificial keys such as patient hospital identification number. Furthermore, access to a sequence of images demonstrating the evolutionary transformation of medical objects (e.g., organ development, disease progression, etc.) is required to understand the dynamics of the medical object under investigation.

Finally, multimedia medical database systems have advanced data analysis and visualization requirements. Several institutions are developing specialized algorithms for computer-aided diagnosis of digital images (e.g., nodule detection in chest x-rays) as well as new image visualization techniques

*This work is supported in part by a grant from the National Science Foundation.

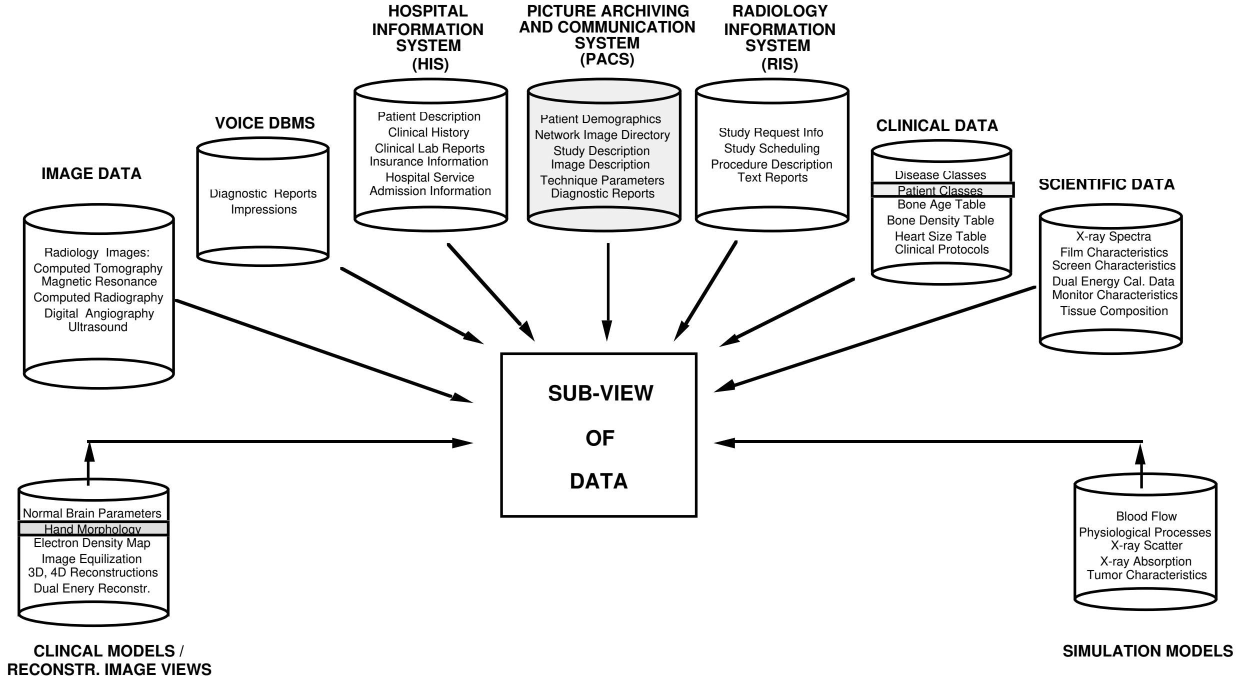

Figure 1: *UCLA Medical Center multimedia database federation. Each of the database systems shown were implemented independently.*

(e.g., 3D imaging, target tissue enhancement, functional mapping, dual energy reconstruction) to improve medical diagnoses. These image analysis techniques generate other types of data characterizing diagnostic image patterns and/or image features related to a disease state. Some examples include cardio-thoracic ratio, fractal shape parameters for brain contours, tumor/organ volumetric data, spatial Fourier spectral information, morphological shape parameters for various bones in the hand, segmentation information for various organ/structural boundaries, and critical gray level breakpoints used to optimize tissue visualization. It is desirable to incorporate these multimedia data analysis and visualization techniques as an integral part of the query processing operation.

2 PROJECT GOALS

Our first goal is to integrate the major medical information systems within our Medical Center: HIS, RIS, and PACS. This goal is driven by the development of a national scientific database system for breast cancer research. The research requires patient history, family history, pathology, and medical workup records from the Medical Center's HIS system, diagnostic reports from the RIS system, and mammographic images from the PACS system. Figure 2 shows the network configuration of these systems.

The goal to access images by content and collections of images by evolutionary processes is being addressed by the UCLA Knowledge-based Multimedia Medical Distributed Database System (KMeD) project. KMeD addresses the data modeling of both pictorial and non-pictorial data in both the spatial and temporal domains and the development of a high-level user query interface. We have initially focused on developing a database system for musculo-skeletal research (with future extensions to brain modeling research). Our goals are to develop an intelligent database system from which researchers can specify: a) query predicates involving bone structure, growth, and evolution, b) analysis methods for feature extraction and characterization, c) visualization methods for query results.

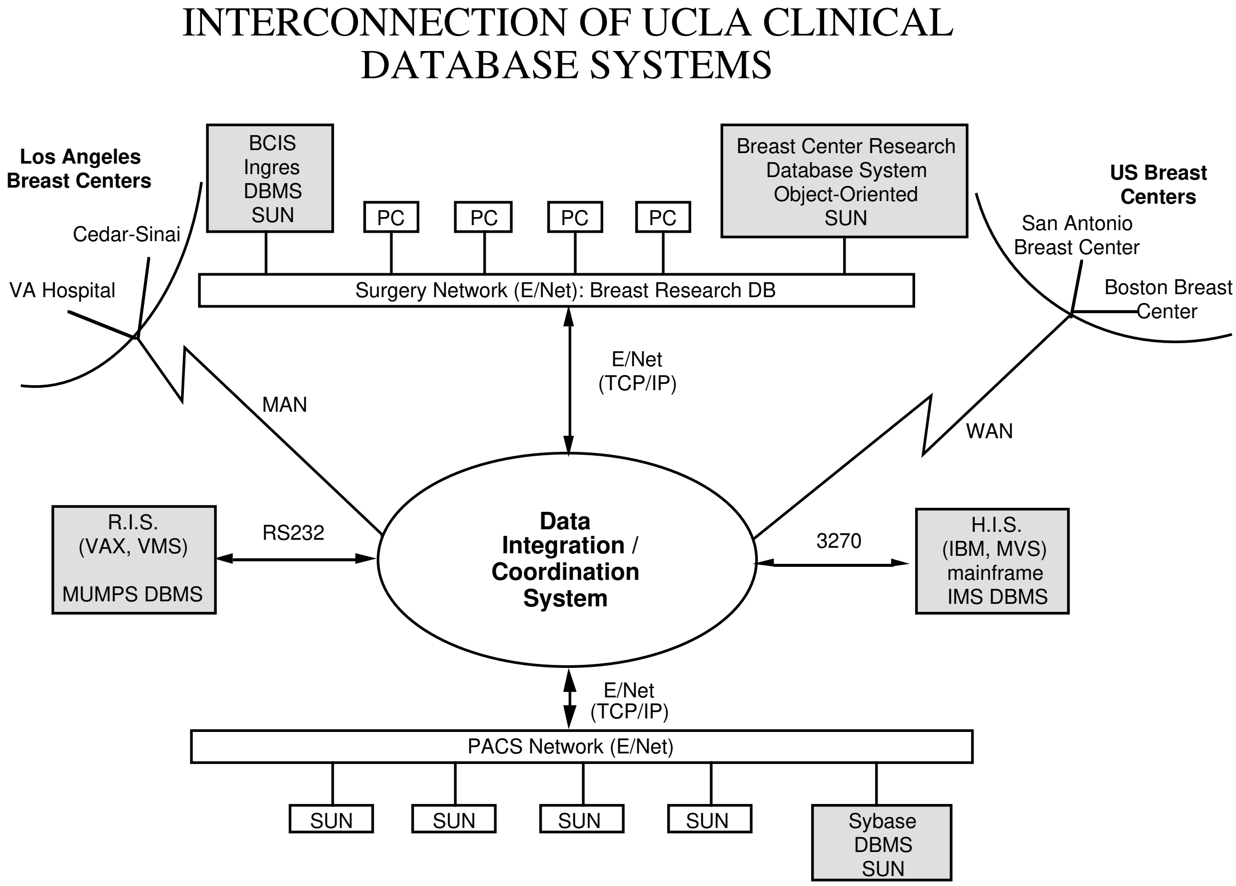

Figure 2: *Network diagram of various databases required for breast cancer research.*

3 DATA INTEGRATION, MODELING, AND QUERYING

The integration of the conventional HIS/RIS/PACS databases required for the breast cancer research project is being performed using the commercially available product Integration Works (Data Integration Solutions Corp.). The system provides high-level CASE tools (collectively called the "Integration Consultant") for designing, modeling, and testing the integration specifications. The CASE tools execute in an object-oriented development environment and include graphical interfaces for global schema design, definition of data integrity constraints, setup of communication protocols, and definitions for source and target machine data translations. The results of the integration consultant are a collection of communication scripts and data translators that are compiled and embedded into the run-time data integration server process (called the "Integration Coordinator"). Applications execute job managers running under an X-Windows environment. Job managers run the automatically generated communication and translation scripts.

The development of improved data models and advanced query processing capabilities for multimedia medical databases is addressed by the KMeD project. Accurate data models are of great interest in medical databases due to the complexity of object features and the complexity of object relationships. The need to model the evolution and dynamics of objects in the body lead to KMeD's temporal evolutionary data model. The requirement for storing, analyzing, and visualizing image data lead to KMeD's stacked image data model.

The extension of traditional object-oriented data models into the temporal domain is an important requirement for accurately representing the data stored in medical image databases. This is because structures in the human body are not static and often change their characteristics and/or existence over time. Our temporal evolutionary data model (TEDM) handles complex queries involving the comparison of the lifespan of two objects and/or events, the creation of objects, the fusion of objects, the fission of an object, and the gross transformation of object properties [Chu92a]. Formal constructs for the temporal relationships between objects, the specification of temporal constraints, and the inheritance characteristics passed between objects have been developed. TEDM uses evolutionary networks for modeling various object transformation processes and includes inheritance rules between objects that exist in various time domains. We have developed TEDM models capturing the development phases of the maturing human hand. The model includes, for example, representation of the closure (or fusion) of the phalange diaphysis and epiphysis hand objects. This model allows us to store and retrieve image collections based on temporal and evolutionary processes.

Image data is modeled using a dynamic stacked logical data structure [Josep88]. Stacks consist of two-dimensional variables or pictures registered to the same gridded coordinate system. Multiple stacks of pictures can provide, for example, additional temporal, spectral, and/or spatial information. Stacks may consist of a hierarchy of other stacks. Furthermore, pictures composed of objects with pictorial as well as non-pictorial attributes and relationships can be represented. Thus, objects at different levels of abstraction can be modeled including time varying image stacks. KMeD users may view an image stack at a high level while still maintaining the flexibility to manipulate images with pixel-level granularity.

The high-level domain-independent query language PICQUERY+ has been designed, and a subset of it is being implemented for the KMeD project [Carde93]. Queries are specified using either a tabular user interface or in a more application specific user friendly graphical interface. Both interfaces allow the user to specify high level operations including query definition (including temporal and evolutionary predicates), data analysis methods, and visualization methods for results. PICQUERY+ also provides menu options for visualizing knowledge including various type abstraction hierarchies [Chu92b] and rules as the user defines the query. Using PICQUERY+ we are able to conveniently express, for example, the following high level medical imaging queries:

- *Show brain cases which demonstrate abnormally shaped ventricles.*
- *Find an image of the proximal phalanx of the fifth finger for patient A.J. Smith and obtain the length of the major axis of this bone.*
- *Retrieve objects in image P2 similar in shape to brain objects considered abnormal in image P1.*

4 CURRENT PROJECT STATUS

The first phase of the integration of the RIS system to a general-purpose networked computing environment is complete. Access to RIS records is accomplished using a more friendly graphical X-Windows user interface. No modifications and/or disruptions to the RIS system were required. The integration server process appears as simply another user to the RIS system. Development time using the integration CASE tools was approximately 2 person days after user requirements and schema definitions were provided. Integration of the HIS and PACS systems are currently in progress.

We are applying the KMeD concepts discussed herein (TEDM, stacked data model, PICQUERY+) to a musculo-skeletal database containing digitized hand radiographs. Custom PICQUERY+ graphical user interface screens are being developed for musculo-skeletal researchers to allow investigators to specify patient demographic profiles, disease pathology constraints, image feature profiles, and evolutionary event constraints. Our preliminary experience suggests that our approach is feasible for retrieving images by content in pictorial databases.

5 REFERENCES

[Carde93] A.F. Cardenas, I.T. Ieong, R.K. Taira, R. Barker, C.M. Breant, "The Knowledge-Based Object-Oriented PICQUERY+ Language", *IEEE Transactions on Knowledge and Data Engineering*, August 1993.

[Chu92a] W.W. Chu, I.T. Ieong, R.K. Taira, C.M. Breant, "A Temporal Evolutionary Object-Oriented Data Model and Its Query Language for Medical Image Management", *Proceedings of the Very Large Database (VLDB) Conference*, Vancouver, Canada, 1992.

[Chu92b] W.W. Chu and Q. Chen, "Neighborhood and Associative Query Answering", *J of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies*, pp. 355–382, Dec. 1992.

[DataI93] Integration Works: Integration Coordinator User's Guide, Data Integration Solutions Corp., 1993.

[Josep88] T. Joseph and A.F. Cardenas, "PICQUERY: A High Level Query Language for Pictorial Database Management", *IEEE Transactions on Software Engineering*, vol. 14, no. 5, pp. 630–638, 1988.