AD_____

Award Number: DAMD17-00-1-0193

TITLE: Digital Breast Imaging Warehouse for Research and Clinical Decision Support

PRINCIPAL INVESTIGATOR: Hong Zhang, Ph.D.

CONTRACTING ORGANIZATION: University of California at San Francisco
San Francisco, California 94143-0962

REPORT DATE: April 2001

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2001	Annual Summary (13 Mar 00 - 12 Mar 01)

4. TITLE AND SUBTITLE
Digital Breast Imaging Warehouse for Research and Clinical Decision Support

5. FUNDING NUMBERS
DAMD17-00-1-0193

6. AUTHOR(S)
Hong Zhang, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California at San Francisco
San Francisco, California 94143-0962

E-Mail: mzh21@yahoo.com

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Breast imaging is used intensively for breast cancer detection. As routine screening examination becomes more popular for women over 40, tremendous amount of breast imaging data has been accumulated. To diagnose breast abnormalities effectively, radiologists expect convenient access to well-organized breast imaging related information. Most breast cancer related data, however, spread across different systems, making them hard to access.

For this project, the investigator will design and build an infra-structural information system by incorporating various kinds of breast imaging data, including patient demographics, family history, radiological reports, digital mammography, ultrasound image and MRI, from a diversity of existing clinical systems, into a digital warehouse. Both clinical- and research-oriented applications will be developed to explore the voluminous amount of data collected in the warehouse. The proposed applications include: (1) case finding, and (2) data mining. Case finding capability virtually transforms the warehouse into a digital case library, and facilitates continuous education and on-the-job decision support, by providing exemplary study to compare with case in question. Data mining application demonstrates the value of consolidating enormous amount of data into a centralized data repository for clinical analysis and decision support. The proposed breast imaging warehouse itself, in the long run, will serve as a reusable platform to foster further breast cancer research and clinical decision making.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast imaging, digital mammography, breast cancer, information system, data warehouse, database	16
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Annual Summary

1. Introduction

Breast cancer is one of the most commonly diagnosed cancers among women in the United States. Approximately 176,300 new cases were diagnosed in 1999 and an estimated 43,000 deaths from breast cancer occurred, making breast cancer the second leading cause of cancer deaths in women after lung cancer [Landis, et. al., 1999].

The early detection of breast cancer increases the survival rate in women [Rickard, 1999]. If detected in the early stages, breast cancer is highly treatable by surgery, radiation therapy, chemotherapy, and hormonal therapy. Combined data from eight randomized studies show that women from the ages of 40-49 entering mammography screening trials had a reduced mortality rate of 17% after 15 years [NIH Consensus Development Panel, 1997]. For women in ages of 50-69 who had their screening done, the mortality rate was even more greatly reduced. Therefore, mammography is an extremely important imaging modality for early detection of breast cancer.

As routine screening examination becomes more popular, tremendous amount of breast imaging data has been accumulated. To diagnose breast abnormalities effectively, radiologists expect convenient access to well-organized breast imaging related information. Most breast cancer related data, however, spread across different systems, making them hard to access.

Furthermore, clinical information systems, such as Hospital Information System (HIS), Radiological Information System (RIS) and Picture Archiving and Communication System (PACS), usually have strict operation performance requirements and high consistent usage. By contrast, decision support systems typically have varying performance requirements, unpredictable workloads, large units of work, and erratic utilization. These differences can make it very difficult to combine clinical operational support and decision support processing within a single information system, especially with respect to capacity planning, resource management, and system performance tuning. For these reasons, system administrators are usually reluctant to allow decision support activities performed on their clinical systems. Decision support data usually needs to be collected from a variety of operational systems (often disparate systems) and kept in a centralized data store resided on a separate platform. That separate data store is a data warehouse. The term of data warehouse originated in the late 1980s, though the concept is somewhat older [Devlin 1988, Inmon 1988]. W.H. Inmon defines a data warehouse as "a subject-oriented, integrated, nonvolatile, time-variant data store in support of' management's decisions" [Inmon 1992]. Data warehouses arose for two reasons: First, the need to provide a single, clean, consistent source of data for decision support purposes; second, the need to do so without impacting operational systems.

An integrated digital mammography data warehouse can facilitate diagnosis, education, and research in the area of breast cancer. Being able to consolidate clinical information and organize

patient files by their images, family histories, pathologies, and demographics can provide useful information for both a clinician and a researcher to diagnose, learn, and study various aspects of breast cancer. Therefore, we proposed to construct a digital data warehouse that can provide important information for both clinicians and researchers at UCSF.

2. Body

The purpose of this study is to build an integrated breast imaging data warehouse to facilitate both breast cancer research and clinical practice. The hypothesis of the project is that diverse breast cancer information, such as patient demographics, family history, medical images of different modalities, and radiological reports, can be consolidated into a digital warehouse, which will serve as a platform to foster research and clinical decision making on breast cancer.

Two major tasks were proposed for Year 1. Task 1 (month 1-3) was completed as planned. Task 2 (month 4-12) was partially completed, due to the execution difficulty described below. Following are the discussion of the specific accomplished tasks as outlined in the approved Statement of Work.

Task 1. Analyze system requirements and design system architecture (Months: 1-3)

The investigators planed to design and build an infra-structural decision support system by incorporating various kinds of breast imaging data, from a diversity of clinical information systems, into a digital breast imaging warehouse. Both research- and clinical-oriented applications will be developed to take advantage of the voluminous amount of data collected in the warehouse.

We have completed the prilimary work of designing the system architecture, including the system modules, data model, interconnection methods, and networking environments. The system architecture of the data warehouse system consists of three layers: (a) the consolidated warehouse, (b) the clinical and research applications, and (c) the front-end user interface. All layers will be built on top of existing clinical systems (data sources) at UCSF Medical Center. A multi-tier architecture has the advantage of separating different computational logic while maintaining necessary cooperation between tiers.

A data model compliant to the American College of Radiology BI-RADS standard [ACR, 1998] was designed to capsulate critical radiological attributes of mammography examinations, which include findings (mass, calcification, architectural distortion, associated findings), breast composition, assessment categories, overall impression, etc. A database system that reflects the designed schema was implemented on a Sun UltraSPARC20 workstation running Oracle Database Management System (DBMS) version 8.0.4.

Detailed description of the architecture, system modules and networking environments can be found in the paper presented to SPIE Medical Imaging 2001 Conference [Zhang, 2001].

Task 2. Develop data acquisition toolkit and collect breast imaging cases: (Months: 4-12)

DICOM [ACR-NEMA, 1998] compliant query/retrieve software modules were developed over MIR Central Test Node Library [Moore, 1994]. These modules support standard DICOM services such as C-Echo, C-store, C-Find, and C-Move. The query and retrieve of digital images can be performed based on patient name, patient ID, and/or study number. By using the implemented modules, we can import the patient information and images from the UCSF digital mammography archive either interactively or automatically, running at night when clinical traffic is low.

A collective set of about 869 cases that were accumulated during a pilot telemammography project conducted at the Laboratory for Radiological Informatics [Lou, 1997, 1998]. With the assistance of radiologists from Mt. Zion. (the Breast Imaging Section of UC San Francisco Medical Center), 434 of the patient cases were categorized. The categorization was based on inclusion of a full spectrum of mammographic findings, which are encountered in clinical practice and defined in the BI-RADS standard. We have identified 170 calcification cases, 115 mass cases, 53 no-finding cases, and 96 undetermined cases. We had planned that selected exemplary cases and associated information from the digital mammography archive and the Radiological Information System (RIS) will be gradually imported into the data warehouse system during the implementation phases.

The study, however, was paused after about six months of execution because of a major administrative difficulty encountered - two key technical investigators/supervisors in the original proposal left the Univ. of California, San Francisco during the first half of year 2000. Both of them are accomplished researchers who have special background, unique training and decades of experience in medical imaging and informatics that are very hard to find replacement for. In absence of the mentorship, it is virtually impossible to achieve the challenging research goals within the proposed time frame. Therefore, the PI decided to return this pre-doctoral training award to the sponsoring agency so that the resources can be re-allocated.

One paper based on the work conducted was presented and published in the 2001 SPIE Medical Imaging Conference.

3. Key Research Accomplishments

- Finished design of preliminary system architecture.
- Designed a data model to encapsulate the attributes/lexicon defined in ACR BI-RADS.

- Implemented a prototype digital mammography database running in Oracle8.
- Developed DICOM-compliant Query/Retrieve software modules to query and retrieve digital mammography images from PACS archive.
- Reviewed and categorized 434 patient cases based on radiological findings, pending import into the data warehouse.

4. Reportable Outcomes

- SPIE 2001 Proceeding paper
- Poster presentation at SPIE Medical Imaging 2001 Conference
- M.S. Degree in Medical Information Sciences conferred by UC San Francisco

5. Conclusions

The purpose of this study is to build an integrated breast imaging data warehouse to facilitate both breast cancer research and clinical practice. We have completed the system design phase, designed a ACR BI-RADS compliant data model, built a preliminary database system, developed DICOM software modules to query/retrieve patient images from central archive and categorized 434 patient cases based on mammographic findings. The project was discontinued after six month of execution due to an unexpected administrative difficulty.

References

S.H. Landis, T. Murray, S. Bolden, and P.A. Wingo, "Cancer statistics, 1999," *Ca: a Cancer Journal for Clinicians*, vol. 49, pp. 8-31, 1, 1999.

M.T. Rickard, "Current issues in mammographic breast cancer screening," *Hospital Medicine*, vol. 60, pp. 325-8, 1999.

National Institutes of Health Consensus Development Panel: National Institutes of Health Consensus Development Conference Statement: breast cancer screening for women ages 40-49, January 21-23, 1997. *J. Natl Cancer Inst* 89(14): 1015-1026, 1997

B.A. Devlin BA and P.T. Murphy, An architecture for a business and information system *IBM Sys. J.* 27, No. 1, 1988

W.H. Inmon, Data Architecture: The Information Paradigm Wellesley, Mass., QED Information Sciences, 1988

ACR, *Breast Imaging Reporting and Data System*, 3rd ed. Reston, VA, 1998.

ACR-NEMA. DICOM: Digital Image Communication in Medicine, 1998

ACR-NEMA. DICOM: Digital Image Communication in Medicine, 1998

S.M. Moore, S.A. Hoffman, D.E. Beecher, "DICOM Shareware: A Public Implementation of the DICOM Standard," in Medical Imaging 1994-PACS: Design and Evaluation, R. Gilbert Jost, Editor, Proc SPIE 2165, pp. 772-781, 1994

S.L. Lou, E.A. Sickles, and H.K. Huang, "Full Field Direct Digital Telemammography: Technical Components, Study Protocols, and Preliminary Results," *IEEE Trans. Info. Tech. in Biomed.*, vol. 1, pp. 31-40, 1997.

S.L. Lou, H.K. Huang, and E.A. Sickles, et. al. "Full-Field Direct Digital Telemammography - System Implementation," SPIE Medical Imaging - PACS Design and Evaluation, vol. 3339, pp.156-164, 1998.

H. Zhang, X. Cao, S.T.C. Wong, S.L. Lou, E.A. Sickles, "Developing a digital mammography data warehouse", SPIE Medical Imaging 2001: PACS and Integrated Medical Information Systems: Design and Evaluation, vol. 4323 (in press), 2001

Appendix

1. Manuscript: Developing a digital mammography data warehouse, H. Zhang, X. Cao, S.T.C. Wong, S.L. Lou, E.A. Sickles, SPIE Medical Imaging 2001: PACS and Integrated Medical Information Systems: Design and Evaluation, vol. 4323 (in press)

Developing a digital mammography data warehouse

H. Zhang, X.H. Cao, S.T.C. Wong, S.L. Lou, E.A. Sickles*

Laboratory for Radiological Informatics,
University of California, San Francisco

**Breast Imaging Section, Mt. Zion Hospital,*
University of California, San Francisco

ABSTRACT

Early detection of breast cancer is believed to be the best means to reduce mortality. Mammography is used intensively for breast cancer detection. As routine screening examination becomes more popular, tremendous amount of breast imaging data has been accumulated. To diagnose breast abnormalities effectively, radiologists expect convenient access to well-organized breast imaging related information. Most breast cancer related data, however, spread across different information systems and represented in incoherent format, making them hard to access on line.

This paper discusses our initial efforts to design and develop a digital mammography data warehouse to facilitate clinical and research activities. Data warehouse is a complete and consistent integration of data from many information sources. It enables users to explore the warehouse for various analysis and decision support purposes. We are designing an infra-structural information system by incorporating various kinds of breast imaging data, from a diversity of existing clinical systems, into a digital data warehouse. Various types of breast imaging data, including patient demographics, family history, digital mammography and radiological reports, will be acquired from the University of California San Francisco (UCSF) digital mammography PACS modules, as well as Radiological Information System (RIS).

Clinical applications are being developed to explore and analyze the voluminous amount of data collected in the warehouse. Case finding, one of the scenarios being implemented, empowers the clinical users with the capability to search the warehouse using a rich set of interested attributes, such as family history, assessment category, and mammographic findings, e.g., mass, calcification, associated findings; shape; distribution; etc.

Keywords: Digital mammography, breast imaging, data warehouse, clinical information system

1. BACKGROUND

Breast cancer is one of the most commonly diagnosed cancers among women in the United States. Approximately 176,300 new cases were diagnosed in 1999 and an estimated 43,000 deaths from breast cancer occurred, making breast cancer the second leading cause of cancer deaths in women after lung cancer [Landis, et. al., 1999].

The early detection of breast cancer increases the survival rate in women [Rickard, 1999]. If detected in the early stages, breast cancer is highly treatable by surgery, radiation therapy, chemotherapy, and hormonal therapy. Combined data from eight randomized studies show that women from the ages of 40-49 entering mammography screening trials had a reduced mortality rate of 17% after 15 years [NIH Consensus Development Panel, 1997]. For women in ages of 50-69 who had their screening done, the mortality rate was even more greatly reduced. Therefore, mammography is an extremely important imaging modality for early detection of breast cancer.

As routine screening examination becomes more popular, tremendous amount of breast imaging data has been accumulated. To diagnose breast abnormalities effectively, radiologists expect convenient access to well-organized breast imaging related information. Most breast cancer related data, however, spread across different systems, making them hard to access.

Furthermore, clinical information systems, such as Hospital Information System (HIS), Radiological Information System (RIS) and Picture Archiving and Communication System (PACS), usually have strict operation performance requirements and high consistent usage. By contrast, decision support systems typically have varying performance requirements,

unpredictable workloads, large units of work, and erratic utilization. These differences can make it very difficult to combine clinical operational support and decision support processing within a single information system, especially with respect to capacity planning, resource management, and system performance tuning. For these reasons, system administrators are usually reluctant to allow decision support activities performed on their clinical systems. Decision support data usually needs to be collected from a variety of operational systems (often disparate systems) and kept in a centralized data store resided on a separate platform. That separate data store is a data warehouse. The term of data warehouse originated in the late 1980s, though the concept is somewhat older [Devlin 1988, Inmon 1988]. W.H. Inmon defines a data warehouse as "a subject-oriented, integrated, nonvolatile, time-variant data store in support of management's decisions" [Inmon 1992]. Data warehouses arose for two reasons: First, the need to provide a single, clean, consistent source of data for decision support purposes; second, the need to do so without impacting operational systems.

An integrated digital mammography data warehouse can facilitate diagnosis, education, and research in the area of breast cancer. Being able to consolidate clinical information and organize patient files by their images, family histories, pathologies, and demographics can provide useful information for both a clinician and a researcher to diagnose, learn, and study various aspects of breast cancer. Therefore, we proposed to construct a digital data warehouse that can provide important information for both clinicians and researchers at UCSF. This data warehouse extends the systems architecture and research results of the UCSF multimodal neuroimaging data warehouse project [Wong 1999].

2. DESIGN OBJECTIVE

Data warehouse is a complete and consistent integration of data from many sources. It enables user to explore the warehouse for various decision support purposes. The objective of this project is to design and build an information infrastructure by incorporating various kinds of breast imaging data from a diversity of clinical systems into a digital warehouse. Specifically, to investigate the hypothesis, the overall goal of the project is twofold: (1) design and implementation of a breast imaging data warehouse; (2) development of application packages for clinical user and researchers to take advantage of the vast amount of data integrated. To explore and analyze the voluminous amount of data collected in the warehouse, case finding application will be developed. It empowers the clinical users with the capability to search the warehouse using a rich set of interested attributes, such as family history, assessment category, and mammographic findings, e.g., mass, calcification, associated findings; shape; distribution; etc. The data warehouse will be equipped with an easy-to-use web-based graphical user interface. In addition, a standardized interfacing is required between the data warehouse and many different databases such as HIS and RIS, using industry standards such as DICOM and HL7.

The project will provide: (1) a reusable breast imaging data model and the methodology to construct the breast imaging warehouse; (2) integrated case search and visualization tools to better serve the radiologist community; and (3) the digital warehouse itself, in the long run, to serve as a reusable platform to foster further education, research and diagnosis in the area of breast cancer. The result should be a versatile data warehouse that can be easily accessed by anyone with appropriate authorization. Moreover, the data warehouse will be a source that promotes more extensive and expansive analyses into the disease of breast cancer.

3. MATERIALS AND METHODS

3.1 System Architecture

Figure 1 illustrates the overall architecture of the proposed data warehouse system. It basically consists of three layers: (a) the consolidated warehouse, (b) the clinical and research applications, and (c) the front-end user interface. All layers are built on top of existing clinical systems (data sources) at UCSF medical center. A multi-tier architecture has the advantage of separating different computational logic while maintaining necessary cooperation between tiers [Fowler, 1996].

Breast imaging data will be collected from several existing clinical information systems across the networks at UCSF, including HIS, RIS and PACS. Notably, a pilot telemammography project conducted at the Laboratory for Radiological Informatics (LRI) has accumulated over 800 digital mammography cases, providing an essential data source for this proposed project [Lou 1997; Lou 1998; Huang 1999].

These acquired data will be cleansed, transformed, and aggregated into a single data warehouse according to an object-oriented data model under development, conforming to the ACR (American College of Radiology) defined BI-RADS (Breast

Imaging Report And Data System) standard [ACR 1998]. Critical breast imaging entities, e.g., mass, calcification, architectural distortion, associated findings, assessment, etc., will be encoded and built into an object-oriented data model. Data from different clinical systems usually carry different formats and cannot be fed into the warehouse directly. To enable fast, reliable and consistent data loading into the warehouse, a full set of data acquisition tools will be developed or integrated to clean, transform, and aggregate original data [Devlin 1997; Friedman, et.al. 1998]. Healthcare communication standards, such as HL7 for clinical textual data, and DICOM for medical image, will be used to communicate with the existing data sources [Oosterwijk 1998; Schadow 1998; Lei 1998].

At the application layer, we will implement algorithms for users to employ in their clinical research or practice. For case finding to support clinical decision making, the proposed approach will encode and index clinical diagnosis extracted from radiological reports and DICOM medical image [Zhang 1998]. The data warehouse will be evaluated in terms of the accuracy, efficiency and completeness to retrieve archived cases from the warehouse according to user defined search.

Besides, an interface based on standardized technology is going to be maintained in order to share the data resource of the system to the public who may be interested in breast imaging related research activities and does not have enough source of breast imaging and/or other related data. The front-end user interface will provide integrated access to all of the information contained in the warehouse and display the information according to user's preferences and in a user-friendly manner.

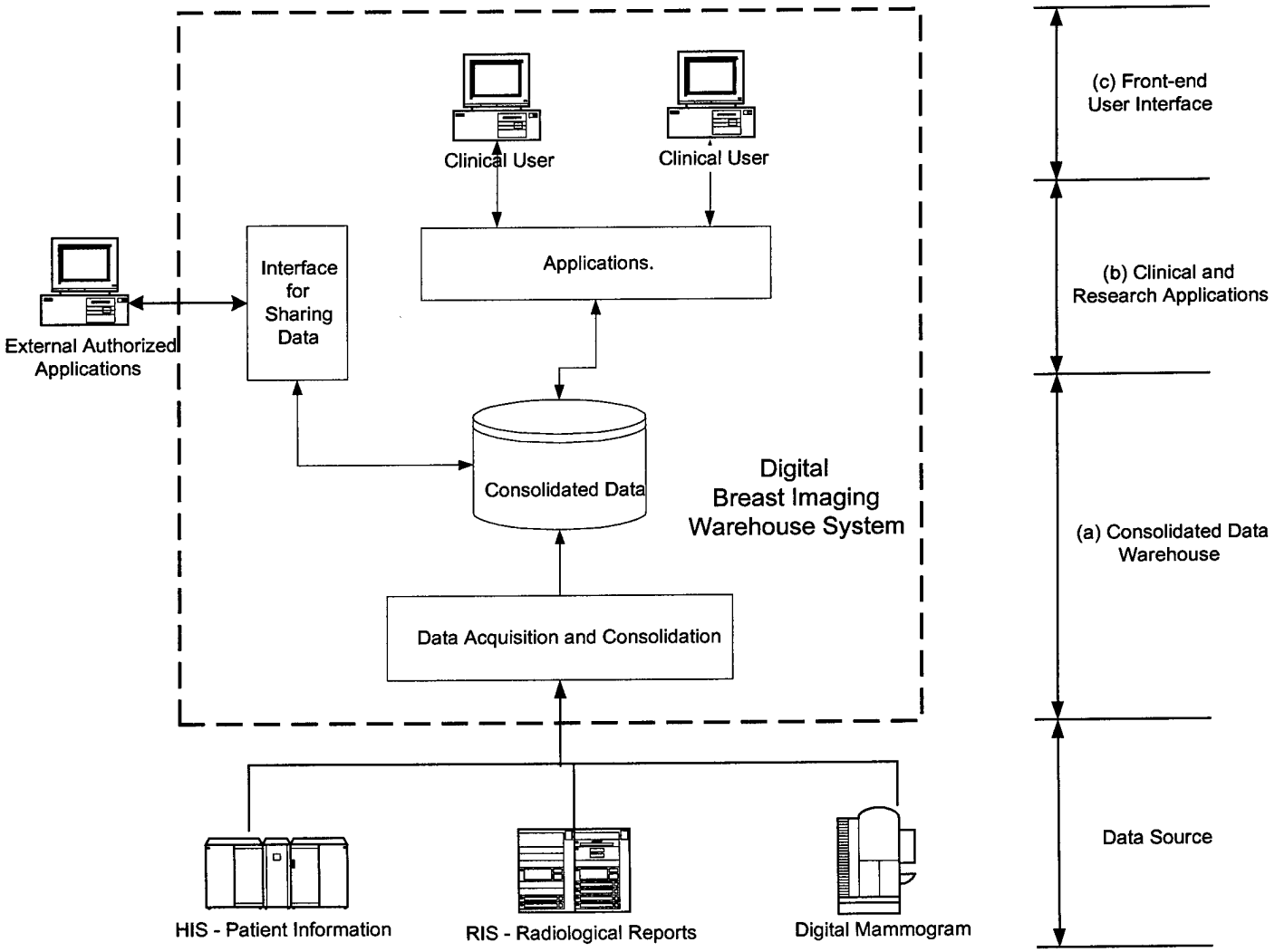

Figure 1 - Layered architecture of the digital mammography warehouse system

3.2 Data Warehouse

For the proposed project, the investigators will be collecting data from various existing clinical operational data systems across the UCSF Medical Center. Multi-dimensional breast imaging data, including patient demographics, related patient history, radiological reports, digital mammography, ultrasound image, MRI, pathology and cytology data, will be acquired from UCSF HIS, RIS, PACS and STOR system:

(1) Textual data

- Patient demographics from UCSF Hospital Information System.
- Diagnostic breast imaging reports from the UCSF RIS system (IDXrad), which will be the major source of radiological findings.
- Related patient history data collected from the paper based medical records at the Breast Imaging Section of the UCSF Radiology department.
- Pathology and cytology report for each core biopsy from the UCSF CDS/STOR system.

(2) Imaging data

- Selected mammogram films archived in the Breast Imaging Section will be scanned using Lumisys Laser Scanner and stored in DICOM format.
- Digital mammogram, which are already in DICOM format and stored in PACS archive, will be solicited into the proposed warehouse.

Many of the issues surrounding data warehousing concern the tasks of obtaining and preparing the data in the first place. The data must be extracted from various sources, cleansed, transported and consolidated, then, loaded into the decision support database. They are also periodically refreshed.

(1) Data extraction is the process of capturing data from operational databases and other sources. Many tools are available to help in this task, including system-provided utilities, custom extract programs, and commercial extract products. Ad hoc program utilities will be developed in house to extract clinical data when off-the-shelf packages are not available.

(2) Few data sources control data quality adequately. As a result, data often requires cleansing before it can be entered into the decision support database. Cleansing of clinical data such as patient demographics will include filling in missing values, correcting typographical and other data entry errors, establishing standard abbreviations and formats, replacing synonyms by standard identifiers, and so on. Data that is known to be in error and cannot be cleansed will be rejected.

(3) Even after it has been cleansed, the data will probably still not be in the form the decision support system requires, and so will need to be transformed appropriately. Usually the required form will be a set of files, one for each table identified in the physical schema; as a result, transforming, the data might involve splitting and/or combining source records along the lines.

(4) Consolidation is particularly important when several data sources need to be merged. In such a case, any implicit relationships among data from distinct sources need to be made explicit.

(5) After finishing all the above data preparation process, data should be loaded, which include (a) moving the transformed and consolidated data into the decision support database, (b) checking it for consistency, and (c) building any necessary indexes.

Detailed processes and data flow among them are illustrated in Figure 2. Breast imaging related data types, including patient demographics, diagnostic findings, related patient history and radiological images are acquired from HIS, RIS, paper records and PACS, respectively. Standard interfaces are applied whenever possible to extract interested information from existing clinical systems. Those established communication standards include HL7 [Beeler 1998] for textual messages and DICOM [ACR-NEMA, 1998] for digital images.

Before insertion of data from various clinical systems into the data warehouse, many steps of pre-processing must be performed, as the diagram in Figure 2 shows. Textual information, after retrieved from HL7 Messenger, will be semantically analyzed. Extraction of radiological findings will be performed on radiological reports, usually in free-text format. These data are then cleaned, transformed, and consolidated into the warehouse. Breast images, either scanner or originally digital, will be

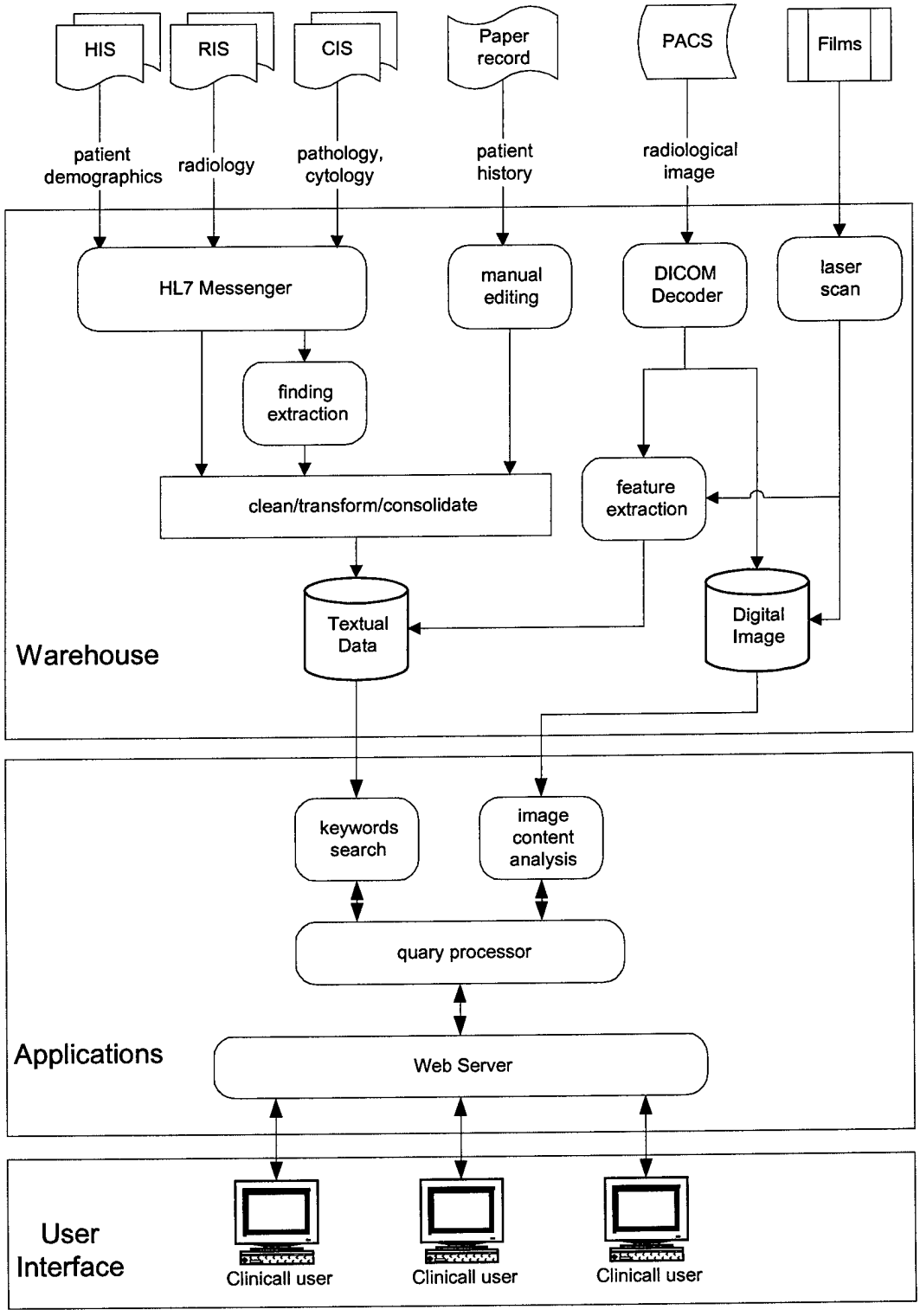

Figure 2. Software model and data flow

analyzed to extract significant image contents. The content descriptors are stored in the textual database, while the images themselves are archived in the digital image database.

As of the application aspect, both clinical and research-oriented applications are web based. For integrated access to the data warehouse, a query processor receives search request from web and parse the request, initiate keyword search and image content search, then consolidate both search results based on user selected criteria.

3.3 Terminology Standard

Mammography reports are often ambiguous and interpretation is indecisive. This is caused by the lack of a universally accepted set of descriptive terms and a structured, decision-oriented reporting system. In order to achieve accurate formulation and communication of the mammography interpretation, this system will be designed to comply with standards for knowledge representation in the domain of breast cancer.

The newly established BI-RADS developed by ACR, NCI, and other institutions is such a standard that we are seeking. It is a quality assurance tool designed to standardize mammography reporting, reduce confusion in breast imaging interpretations, and facilitate outcome monitoring. The key elements of BI-RADS are a lexicon of standardized terminology, a reporting organization and assessment structure, a coding system and a data collection structure. The report organization assists radiologists in providing a succinct review of the mammogram. Results are then communicated to the referring physician in a clear fashion with a final assessment that indicates a specific course of action. Results are compiled in a standardized manner that permits the maintenance and collection analysis of demographic, mammographic and outcome data. Through a medical audit and outcome monitoring, the system provides important peer review and quality assurance data to improve the quality of patient care.

3.4 Resources Used

UCSF operates a Hospital Integrated PACS (HI-PACS), which emphasizes standardization, open systems connectivity, hierarchical memory management, database integration, and data security [Wong 1996]. X-ray, MR, CT, US and digital mammography examinations are centrally archived in the UCSF HI-PACS, which is the major source for all radiological studies. In addition, both Hospital Information System and Radiological Information System are crucial in the process of integrating patient information.

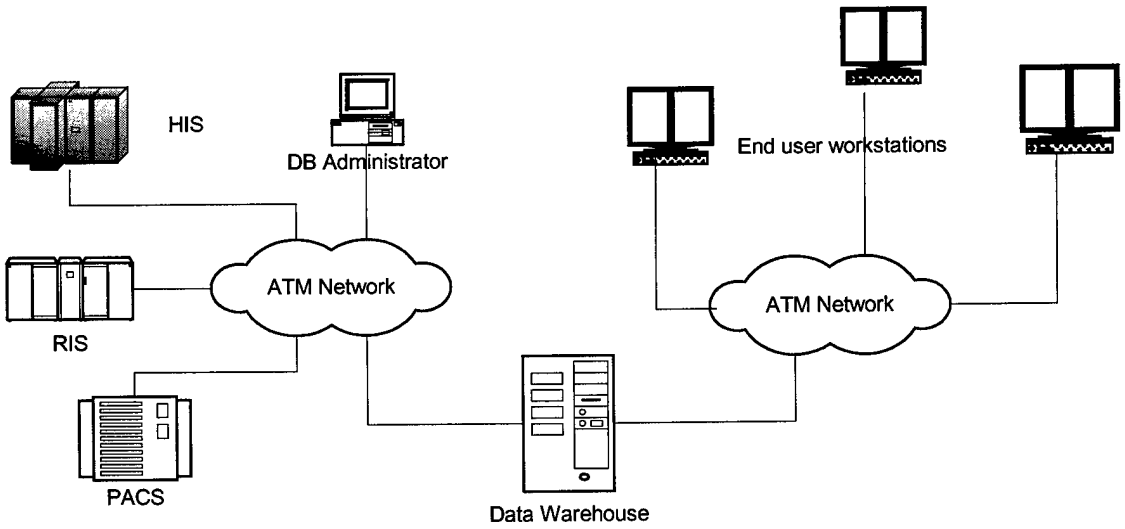

Figure 3. Networking environment

The UCSF HI-PACS networking environments support 622 megabits per second (Mbps) and 155 Mbps ATM links, as well as 10/100 Mbps Ethernet networks. It connects not only directly to various imaging scanners, but also to PACS of subspecialty sections, such as ultrasound, and to radiology, hospital information systems. Image acquisition is based on the DICOM standard while report acquisition relies on the HL 7 standard. All existing clinical systems, the proposed data warehouse, and the end user workstations will be interconnected through the UCSF campus networks. The networking environment is shown in Figure 3.

4. SUMMARY

The vital role of mammography in detecting breast cancer, together with the existing situation that most breast cancer data scattered in various clinical systems prompts us to design and develop a digital mammography data warehouse. The purpose is to develop an integrated information infrastructure by incorporating various kinds of breast imaging data, from a diversity of existing clinical systems, into a digital warehouse. Various types of breast imaging data, including patient demographics, family history, digital mammography and radiological reports, will be consolidated. Clinical applications will also be developed to explore and analyze the voluminous amount of data collected in the warehouse via extensive on-line data query algorithms. Case finding, one of the scenarios being implemented, empowers the clinical users with the capability to search the warehouse using a rich set of interested attributes, such as family history, assessment category, and mammographic findings. The digital mammography data warehouse will be a source and information platform that promotes more extensive and expansive analyses into the disease of breast cancer.

ACKNOWLEDGEMENT

This project is partially supported by the U.S. Army Breast Cancer Research Program (BC990626). The research extends the work of Prof. Wong in the area of Neuroimaging Data Warehouse funded by an NINDS/NIMH R01 Human Brain Program Grant and an NLM R29 FIRST Award. The authors would also like to thank Prof. H.K. Huang of USC for his valuable advice.

REFERENCE:

S. H. Landis, T. Murray, S. Bolden, and P. A. Wingo, "Cancer statistics, 1999," *Ca: a Cancer Journal for Clinicians*, vol. 49, pp. 8-31, 1, 1999.

M. T. Rickard, "Current issues in mammographic breast cancer screening," *Hospital Medicine*, vol. 60, pp. 325-8, 1999.

National Institutes of Health Consensus Development Panel: National Institutes of Health Consensus Development Conference Statement: breast cancer screening for women ages 40-49, January 21-23, 1997. *J. Natl Cancer Inst* 89(14): 1015-1026, 1997

Devlin BA and Murphy PT An architecture for a business and information system *IBM Sys. J.* 27, No. 1, 1988

Inmon WH. Data Architecture: The Information Paradigm Wellesley, Mass., QED Information Sciences, 1988

M. Folwer, *Analysis Patterns: Reusable Object Models*, Addison-Wesley, 1996.

S. L. Lou, E. A. Sickles, and H. K. Huang, "Full Field Direct Digital Telemammography: Technical Components, Study Protocols, and Preliminary Results," *IEEE Trans. Info. Tech. in Biomed.*, vol. 1, pp. 31-40, 1997.

S. L. Lou, H. K. Huang, and E. A. Sickles, et. al. "Full-Field Direct Digital Telemammography - System Implementation," SPIE Medical Imaging - PACS Design and Evaluation, vol. 3339, pp.156-164, 1998.

H. K. Huang and S. L. Lou, "Telemammaography: a technical overview," *RSNA*, 1999. (in press)

ACR, *Breast Imaging Reporting and Data System*, 3rd ed. Reston, VA, 1998.

B. Devlin, *Data warehouse: from architecture to implementation.* Reading, Mass: Addison-Wesley, 1997

C. Friedman, G. Hripcsak, and I. Shablinsky, "An evaluation of natural language processing methodologies," *Proc AMIA Symp*, vol. 174, pp. 855-9, 1998

H. Oosterwijk, "DICOM versus HL7 for modality interfacing," *Journal of Digital Imaging*, vol. 11, pp. 39-41, 1998.

G. Schadow, U. Fohring, and T. Tolxdorff, "Implementing HL7: from the standard's specification to production application," *Methods of Information in Medicine*, vol. 37, pp. 119-123, 1998.

G. P. Lei, H. Zhang, X. Zhou, and A. Wong, "Real-time online operation in DICOM Query/Retrieve software module," *Radiology*, vol. 209(p), Nov. 1998, pp582.

H. Zhang, G. P. Lei, A. Wong, "Portable DICOM-compliant PC/NT-based Diagnostic Workstation with ATM Technology," *Radiology,* vol. 205(p), Nov. 1997, pp308.

Beeler GW. HL7 Version 3 - An object-oriented methodology for collaborative standards development. *International Journal of Medical Informatics* 48: 151-161, 1993

ACR-NEMA. DICOM: Digital Image Communication in Medicine, 1998

S. T. C. Wong and H. K. Huang, "A hospital integrated framework for multimodal image base management." *IEEE Trans. Systems, Man, and Cybernetics*, 26(4), 1996:455-469.

S. T. C. Wong and Donny Tjandra, "A digital library for biomedical imaging on the Internet," *IEEE Communication*, Jan 1999, pp. 84-91.