"Permanent" Records: Experience With Data Migration in Radiology Information System and Picture Archiving and Communication System Replacement

Fred M. Behlen, Richard E. Sayre, James B. Weldy, and Jennifer S. Michael

In the replacement of both a radiology information system (RIS) and a picture archiving and communication system (PACS) archive, data were migrated from the prior system to the new system. We report on the process, the time and resources required, and the fidelity of data transfer. We find that for two PACS archives, both organized according to the Digital Imaging and Communications in Medicine (DICOM) information model, data may be transferred with full fidelity, but the time required for transfer is significant. Transfer from off-line backup media was found to be faster than transfer from our robotic tape library. In contrast, the RIS replacement required extensive labor to translate prior data between dissimilar information models, and some data were inevitably lost in the translation. Standards for RIS information models are needed to promote the migration of data without loss of content.

RETAINING HARD-COPY RECORDS is a matter of storage of physical objects, but keeping permanent electronic records requires the migration of data from old system to new when systems are replaced. In prior work, we have noted this as a strategic planning issue in information systems architectures,[1-3] and have projected the costs and time required for migration of picture archiving and communications system (PACS) data from a system to its successor.[3] Since then, our institution has replaced both its radiology information system (RIS) and its PACS archive, and 14.5 years of RIS data and 2 years of image data were migrated. This report analyzes the cost and time required for the migration process and the fidelity of data transfer.

The RIS data migration involved the transfer of data from one proprietary information model to another. In contrast, the two systems involved in the PACS data migration, while differing in implementation, both followed the Digital Imaging and Communications in Medicine (DICOM) information model.[4] The implementation experience shows the differences between these two approaches, and illustrates the value of standards.

MATERIALS AND METHODS

RIS Replacement

A homegrown RIS was replaced with a commercial system (QuadRIS, ADAC Healthcare Information Systems, Houston, TX). The homegrown RIS (University of Chicago Hospitals RIS [UCH RIS]) was written in the M Programming Language (MUMPS). The UCH RIS was originally descended from the Missouri Automated Radiology System (MARS)[5] and was nearly completely rewritten during the past 14 years. The new RIS is based on thick-client relational database technology. Its vendor specified a flat-file format for data to be imported into the system. Programs were written for the UCH RIS to create files in this import format.

We had elected to migrate all radiology results from the 14.5-year operating history of the UCH RIS, plus other information (patient, staff, and procedures) referenced in those results. This restriction was desirable because the old RIS database contained patients for whom there were no results, as they had film folders at the time the UCH RIS was installed, or they had examinations scheduled and then canceled. In addition, there were staff and procedures in the old RIS that were not referenced in the results, and these were also to be left behind. It was thus anticipated that the historical data would be "cleaned up" in the data migration process. Many other data structures of the old RIS, such as organizational units (flashcard areas, work areas, referring services), scheduling templates, and accounts and access privileges were not candidates for automated transfer, and were manually transferred, often with extensive reorganization, to the new system.

PACS Archive Replacement

Similarly, a one-of-a-kind PACS archive (Martin Marietta, Valley Forge, PA) was replaced with a new commercial archive (Storage Technology, Louisville, CO). Both systems are based on the DICOM information model.

The old archive employs an ADIC (formerly EMASS) AML-J robotic tape library, with four DLT-4000 tape drives, and controlled by AMASS volume management software (all from ADIC, Redmond, WA). The AMASS software, running on a Sun Ultra 1 workstation (Sun Microsystems, Palo Alto, CA), manages the storage of data on the tapes, and presents a file-system interface to applications programs. DICOM services were provided by Mallinckrodt software tools[6] extensively modified by the integration vendor, running on a Sun 1000 server.

The new archive consists of a Storage Technology 9710 robotic tape library with four type 9840 tape drives, controlled by Storage Tek ASM hierarchical storage management software

From the Department of Radiology, The University of Chicago; and Department of Radiology, The University of Chicago Hospitals, Chicago IL.

Address reprint requests to Fred M. Behlen, PhD, Department of Radiology MC2026, The University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637. E-mail: f-behlen@uchicago.edu.

0897-1889/00/1302-1041$10.00/0
doi:10.1053/jdim.2000.6867

running on a Sun Enterprise 450 server. DICOM services on the new archive are provided by Imageon e-CIMS software (Imageon Solutions, Birmingham, AL). The two servers were in different buildings of the main hospital campus, and were connected using 100-Mbps Ethernet links to the hospital's asynchronous transfer mode (ATM) backbone.

The former archive contained computed tomography (CT), magnetic resonance image (MRI), and digital chest data collected over a 2-year period. Images were stored in DICOM files, in the unpackaged Big Endian message format familiar to users of the older DICOM tools. During the years of operation of the archive, CT and MRI image data were routinely saved on backup tapes outside the robotic library, using a DLT-7000 tape drive directly attached to the Sun 1000 server. Image files were stored on these tapes in tape archive (tar) format. These images were transferred to the new archive using Unix shell scripts that copied each tape to disk and transmitted every image instance to the new archive using the Mallinckrodt send_image utility. Images were transferred in an order related to tape position or alphabetical file name, but not directly related to patient or study: the receiving DICOM application at the new archive sorted and indexed the image data according to identifying information in the header of each DICOM image information object.

Digital chest images from two dedicated chest computed radiography units (Models 9501HQ and 9501ES, Fuji Medical Systems) were stored in duplicate volume groups in the robotic library. Images from the 9501ES model also had soft-tissue and bone images for the posterior-anterior projection, approximately doubling the number of images in a typical chest examination. In addition, there were CT and MRI images from a 2-month period, which were in the robotic library and not included in the backup tapes. All of these image files were read from the tape and transferred to the new archive, using specially optimized shell scripts. Multiple sending streams were maintained, with an average of two drives simultaneously sending data to the new archive. It is critically important that the files be read from the tape in the order they appear on the tape, rather than the order they appear in the virtual directory presented by the volume management system, as the penalty for random access to tape files is orders of magnitude lower throughput. Because we were under pressure to retire the old archive before the end of 1999, we elected to transfer these images to the new archive as files, and to import them into the DICOM archive later. To this end, the shell scripts copied each file to a partition in the new archive's hierarchical storage manager, rather than send it to the DICOM front-end using the send_image utility.

RESULTS AND DISCUSSION

The migration of PACS data was logically straightforward and limited primarily by issues of the large quantity of data transferred. The RIS data migration project, on the other hand, was dominated by the tasks of transforming information between different data models.

The transfer of image data proceeded more quickly than expected. The transfer performance obtained is summarized in Table 1. It was apparent that the limitations of accessing the data in the tape-based robotic library were the rate-controlling step, even after optimization for tape position of image files. The DICOM server on the new archive was able to receive, parse and store images at a rate averaging 1.4 images per second; we did not attempt to determine what was the rate-limiting step in this case.

However, the transfer from the robotic library was much slower, even with a comparable amount of data, much smaller number of images, and no overhead of DICOM parsing and database insertion.

Note that these image files have been transferred to the hierarchical storage of the new tape library, but as of press time for these proceedings (January 2000) have not been imported into the archive's DICOM database. The process of importing the transferred image files into the DICOM archive database will move the data within the archive's computer system, but the old archive's equipment has now been decommissioned.

A key result of the PACS data migration experience was that since both archives followed the same information model (DICOM) for organizing images into patient folders, studies, and series, the image data were transferred without loss of content from the original system to its successor.

Table 1. Transfer of Image Data From Old to New PACS Archive

Data source	DLT-7000 backup tapes	AMASS virtual file system on robotic tape library
Image modality	MRI and CT	Computed radiography
Equipment	Sun Sparc 1000E, one DLT-7000 drive; 100 Mbps link over ATM backbone	Sun Ultra 1, ADIC AML-J using an average of 2 DLT-4000 drives, 100 Mbps link over ATM backbone
Method	Dearchive files from DLT-7000, transmit using DICOM send_image, no production load on receiving server	Remote copy (rcp) each file on a volume, DICOM layer not addressed, no production load on receiving server (except infrequent manual requests)
Volume		
Images	2,400,000 images	115,000 images
Bytes	600,000 MByte	750,000 MByte
Time	20 days	6 weeks

Table 2. Data Conversion Statistics for Results Transferred to New RIS

	UCH RIS	New RIS	% Transferred	Not Transferred
Signed	1,962,083	1,905,753	97.1%	56,330
Total	2,214,719	2,061,609	93.1%	153,110
% Signed	88.6%	92.4%		

The same cannot be said for the migration of RIS data. There are no standards for information models in RIS. Indeed, many vendors regard details of their information models to be proprietary secrets. Information transferred from the old to the new RIS therefore required transformation to fit the information model of the receiving system.

These differences included information in the prior system that could not be represented in the new system, such as multiple radiologists contributing to a single report. The differences also included data fields required in the new system for information not always present in the old system, such as the first name of the referring physician. In these cases, the required fields were filled with dummy data that a human observer can readily recognize. Equally troublesome were semantic mismatches in information that could not be easily classified as omissions or additions. For example, the various states of order status are differently defined in the old and the new RIS. We have adopted the new classification scheme, but there is no way to correctly represent the status of past orders according to the definitions prevailing at that time.

The gross statistics for the data conversion process are presented in Table 2. Unsigned entries in the old system include examinations that were canceled, as well as examinations that were performed and their interpretation was canceled, and other exceptional conditions. Many of these cases were deliberately not transferred to the new RIS. Of the signed results in the UCH RIS, 97.1% have been successfully imported into the new system. More than 56,000 results remain in exception queues as of the press time for these proceedings, and it is estimated that another person-month of full-time labor will be required to adequately resolve remaining issues.

The labor requirements for the data migration to the new RIS are difficult to accurately measure, as a systems conversion is often used an opportunity to "clean things up," and the table-building activities of a system installation include a large component of mapping existing procedure into the model of the new system. Viewed this way, total effort for data migration easily exceeded 1 full-time equivalent year, of which approximately 4 full-time months were required for analysis and programming of the machine transfer of results and related data. In contrast, less than 1 person-month was consumed by the PACS data migration.

CONCLUSIONS

Image data acquired as DICOM can be faithfully propagated from a system to its successor. The overhead of transfer to the DICOM archive was less than previously expected, but still represents a significant time component in the transition to a new PACS system. The value of standardized storage in off-line media was demonstrated by our experience with transfer of DICOM image files from backup tapes.

Acceptable performance in transferring data from the robotic tape library is dependent on presenting requests for images in the order they are stored on the tapes. This is information that is not available through a DICOM interface, and thus transfer of archives cannot be done from the outside, interacting only through the DICOM interface.

A more significant challenge is the faithful migration of RIS data. Commercial RIS systems follow no common information model, and some losses in information content are inevitable. Concern about the long-term fidelity of electronic records subjected to multiple sequential migrations is justified.

Significant benefits in the permanence of these electronic records would follow from adoption of standards for modeling this clinical information. Efforts underway in DICOM and Health Level Seven (HL7)[7] standards organizations are promising, but more involvement in standards development by the radiology and information systems communities is clearly indicated.

REFERENCES

1. Behlen FM, Hoffmann KR, MacMahon H: Long term strategies for PACS archive design. Proc SPIE 2711:2-8, 1996

2. Behlen FM: A DICOM document-oriented approach to PACS infrastructure. J Digit Imaging 11:35-38, 1998 (suppl 1)

3. Behlen FM: The document-oriented approach to electronic patient records, in Waegemann CP (ed): Proceeding Toward an Electronic Patient Record '98, vol 3. Newton, MA, Medical Records Institute, 1998, pp 197-201

4. Digital Imaging and Communications in Medicine (DICOM): *http://www.nema.org/nema/dicom/*

5. Lehr JL: Installation of MARS at The University of Chicago, in Proceedings of the Eighth Conference on Computer Applications in Radiology. Reston, VA, American College of Radiology, 1984, pp 35-45

6. Mallinckrodt CTN (Central Test Node) demonstration software, available from *ftp://ftp.rsna.org/pub/dicom-software-images/ctn/*

7. Health Level Seven standards: *http://www.hl7.org*