Management of the Picture Archiving and Communications System Archive at Texas Children's Hospital

Maria Elissa E. Blado

As hospitals convert from conventional film-based imaging to picture archival and communications systems (PACS), methods for managing an enormous library of images must change considerably. While most hospitals are required to retain general, non-mammographic, radiologic images for 7 years beyond the examination date, our pediatric hospital must maintain images until the child's eighteenth birthday, plus the 5-year statute of limitations. Although the physical extent of an electronic archive is tiny compared with a film archive, a long-range strategy is required to ensure that electronic images acquired today can be retrieved and viewed 23 years in the future. Challenges to the long-term stability of the electronic archive include the limited and uncertain shelf life of high-density electronic storage media, the finite maintainability of the electromechanical systems for reading the media, the short product lifetime of software for accessing the images, rapid development of higher density storage products, and the exponential advancement of computer and networking technology that fuels product obsolescence. Since we cannot assure the function of our current archive in two decades, we are committed to a continual process of migration of old electronic image data to newer media and systems. As an early-adopter of PACS technology, Texas Children's Hospital's (TCH) archive management experience is relevant to others. Although not filled to capacity, our first digital archive, based on phase-change write-once-read-many (WORM) technology, was forced into an inactive status by software and hardware changes. Our second set of archives was partially filled with low-density magneto-optical disk (MOD) media, when the drives were upgraded to high density and then filled to capacity. This undesirable situation forced us into shelf management of media. Our third-generation archive is based on a helical tape library with the capacity to contain 7 years of examinations. We will describe the motivation for data migration, limitations in the methods available to perform the migration, and unanticipated benefits of the migration process.

Copyright © 2001 by W.B. Saunders Company

TEXAS CHILDREN'S Hospital (TCH) began investing in a picture archiving and communications system (PACS) in 1991 and has grown to an extensive PACS infrastructure. For image acquisition, our PACS has 10 computed radiography (CR) devices, which include patient-identification devices and quality-control devices for images. Digital interfaces have been installed for four fluoroscopy units plus two portable C-arms; three ultrasound units plus one portable ultrasound unit; two magnetic resonance (MR) imaging systems with two independent consoles for MR cross-sectional imaging, one each for neuroradiology and cardiology. The computed tomography (CT) operator's console sends images directly to the PACS in digital form, while Nuclear Medicine has three cameras connected to the PACS through a single gateway. TCH PACS also has three laser film digitizers and includes a web server that receives all new images and provides them to PC-based browsers. Image display includes 32 review stations, which support 15 clinical services, together with web-based viewers that support another 32 clinical services. TCH PACS also has conventional film processors and laser film printers for hard copies.

Image management involves four network gateways that receive images from the acquisition stations and route them to destinations for display and storage. There are four archive servers at TCH. Although not filled to capacity, our first digital archive, AS2, based on phase-change write-once-read-many (WORM) technology, was forced into an inactive status by software and hardware changes. Each of our second pair of archives, AS1 and AS3, has a 258-platter 4.1-GB magneto-optical disk (MOD) jukebox (ODJ) with 176 GB of redundant array of inexpensive disks (RAID) each. The jukeboxes were initially populated with low-density MOD media, and later the drives were upgraded to high density. When the jukeboxes reached full capacity, we were forced into shelf management of media, which we refer to as "offline volumes." Our third-generation archive, AS4, is based on a helical tape library with the capacity to contain 7 years of examinations. Its archive device is a 9,710 12.5-TB StorageTek (Storage

From the Texas Children's Hospital, Houston, TX.

Address reprint requests to Maria Elissa E. Blado, Edward B. Singleton Diagnostic Imaging Services, Texas Children's Hospital, 6621 Fannin St, MC 2-2521, Houston, TX 77030. E-mail: meblado@texaschildrenshospital.org.

Copyright © 2001 by W.B. Saunders Company

0897-1889/01/1402-1021\$35.00/0

doi:10.1053/jdim.2001.23825

Technology Corp, Louisville, CO) library with six 9840 drives. Image management is coordinated by an Oracle (Redwood Shores, CA) Server, with demographic and exam information provided by a radiology information system (RIS) interface.

ARCHIVE MANAGEMENT

As an early-adopter of PACS technology, TCH's archive management experience is relevant to others. Figure 1 shows the migration scheme for multiple archive servers implemented at the PACS Diagnostic Imaging Department of TCH. All examinations coming from different acquisition stations have been divided

between the two ODJ archive servers, AS1 and AS3, so that images are not duplicated on these servers. It is possible, though, that an examination may be pulled from the local cache of one server to another (ie, for troubleshooting purposes), and it may be neglected from being deleted from the latter server, so that duplication of examinations may occur between the two servers. The local cache of each of these primary servers is then checked daily to ensure there is no duplication between the two archives. Patient information and demographics are also double-checked on a daily basis. In addition, a report is run daily to verify that examinations performed

according to the RIS match those archived in the PACS.

The secondary archive, AS4, has been receiving everything sent by all acquisition stations since June 2000. For FY99, TCH Diagnostic Imaging reported 115,000 examinations at the annual meeting of the Society of Chairmen of Radiology of Children's Hospital (SCORCH), making it the third busiest pediatric imaging department in the United States.

As the AS1 and AS3 jukeboxes reached capacity, older MODs are removed from the ODJ (as off-line volumes) and stored in their jewel cases inside a locked cabinet. Fresh MODs are imported into the jukebox, on an average of 5 per week for each ODJ.

DATA MIGRATION STRATEGY

The storage capacity of the tape library is sufficient to contain all of our images for 5 to 7 years, including duplicate copies of images already archived on MOD. As soon as the tape library was active, we began migrating data from the MOD jukeboxes to accomplish three purposes. First, we needed to eliminate shelf management of MODs by copying them into the tape library where they will be automatically accessible. Second, we needed to improve the efficiency of the MOD jukeboxes, by copying low-density media to tape and replacing low-density MODs with high-density media. Finally, we decided to duplicate every examination that exists on MOD onto tape. Concurrently, we are recording new examinations on both tape and MOD. This allows us to maintain the most recent examinations on tape and on-line in MOD jukeboxes, with older examinations available on tape, or from MODs stored on the shelf, in case of a tape failure.

Migration of data involves physical rearrangement of the examinations among the media. While originally the examinations may have been ordered chronologically, migrated data are interspersed with new examinations. It occurred to us that, of the examinations residing offline on the shelf, the most important ones were those with the most recent examinations, which were most likely to be recalled for comparison with new examinations. Although we wanted to begin the migration process with the most recent off-line examinations, rather than the oldest, the software available to us would not support this strategy.

TCH's PACS vendor, AGFA Corp (Ridgefield Park, NJ) uses software called Scavenger (Mitra Imaging, Waterloo, Ontario, Canada) to perform the data migration. Scavenger is designed to go through each MOD of the ODJ from the lowest volume ID (generally, the lower the volume ID, the older the examinations that are stored in it). It copies each examination in that MOD onto tape in AS4. Note the difference in storage media between the two transfer points.

Scavenger was launched at the end of May 2000 on one of the main archive servers, AS3. Not to launch Scavenger on both main archive servers was a deliberate decision for the purposes of archive management, as well as to avoid the risk of system overload. As of the first week of January 2001, after 8 months of operation, only 434 of 938 volumes have been fully migrated (46%). This slow progress might be attributed to the size of the TCH PACS system, and to the large number of processes going on a daily basis, plus the database issues Scavenger is uncovering along the way.

Scavenger works in the background so that it does not affect current processes in the production PACS. It "wakes up" after a certain amount of time and checks the number of active processes. The software includes a setting for the maximum number of system jobs before Scavenger can do any work (`max_jobs`). If the number of active system jobs is less than the maximum, Scavenger will continue with the migrating process. If not, it will then go back to "sleep."

Scavenger keeps a file of which disks it has finished "scavenging," which are partly scavenged (or being scavenged), and which have not yet started the scavenging process. System administrators review this file daily so that scavenged disks may be removed from the jukebox, labeled, and stored in the cabinet. Empty ODJ slots are then available either for blank MODs or with off-line volumes. Eventually, all MODs in both ODJs and all off-line disks will be scavenged, so that all examinations will exist in two copies: one on tape and one on MOD.

IMPEDIMENTS TO MIGRATION

`Max_jobs` was initially set to a certain level at Scavenger's launch, but since then, that number was increased several times. It currently is set to twice the initial setting. When any changes are made to that number, we have to verify that the

increase does not interfere with the active system's jobs. Scavenger logs are only capable of holding 1.5 days of transactions. On a Monday, Scavenger's activity for the whole weekend cannot readily be determined.

Scavenger migrates data from the oldest disk forward. It began on an ODJ containing 258 MODs, and with approximately 100 off-line volumes. These off-line volumes are the oldest disks that came from both primary archive servers. As Scavenger migrates data, more disks had to be removed from the first ODJ so that the original set of off-line volumes could be put back into the ODJ for Scavenger to work on.

The non-unique numbering scheme applied by PACS for the MODs became a migration issue. An MOD would be identified as 000001 for the first ODJ and a second MOD would be given the same volume ID, namely, 000001, for the other ODJ. The original set of off-line disks had twins for volume IDs. Scavenger reports the volume ID of any completely scavenged disk so that it may be removed. Unless a volume's twin is not also completely scavenged, it is impossible to identify the completely scavenged disk, because Scavenger's report does not include the ODJ slot location of the scavenged disks. The next version of the PACS software corrects this error, issuing unique identification for MOD volumes in multiple ODJ's.

Each ODJ needs fresh MODs to record new examinations. Unfortunately, the number of fresh disks introduced into the jukebox is greater than the number of completely scavenged disks each week. In order to make space for fresh MODs, older disks must be removed as off-line volumes. Scavenger is running only on one archive, but there are two jukeboxes that need fresh disks so twice as many disks are removed and stored off-line.

This creates a situation where we sometimes have a more recent prior examination for a patient in an off-line volume than on-line. Our radiologists recognized that it is more valuable clinically to have the most recent prior on-line than an older prior. This is significant in designing software that performs migration: migration of the most recent examinations is more useful than starting from the oldest disk.

Our oldest archive server, AS2, has been inactive for 4 years. Examinations stored here were performed between early 1992 to early 1997, and the images can be made available upon request.

The manual process of transmitting the examination from AS2 to the active system is straightforward for individual examinations.

We are considering migrating the examinations stored on AS2. The RIS interface was not active until early 1997, so demographic information associated with these examinations has not been verified to the RIS. Migrating the examinations without RIS verification risks corrupting the active database. The current method for importing examinations from AS2 would require a significant commitment of time because this jukebox has nearly 42,000 examinations. Automating the verification process would involve re-broadcasting demographic information from the RIS to our RIS interface.

Data migration imposes overhead on all the PACS core components involved, as well as additional network traffic. To accommodate this increased demand, we are upgrading our core components from Fast Ethernet to Gigabit Ethernet speeds.

DATABASE ISSUES

Scavenger has uncovered discrepancies in our image database and essentially is performing a purge of our database. Otherwise, such issues would have been dealt with incidentally when users call about specific examinations. Scavenger has caused us to address these in their entirety.

One such discovery is that some examinations have their images split between two or more MODs, which causes problems when one of these MODs is an off-line volume. These studies in question are completely archived into our tape library. Our PACS vendor is working on a script to correct all such examinations in the database.

Scavenger sometimes has difficulty locating certain examinations. The database reports that the examination is located in a certain MOD, but when Scavenger searches that MOD, the examination is missing. The database has some mismatches in examination location information that must be corrected manually.

We have also encountered a few MODs that either contain corrupt images or that the ODJ will not accept during an off-line volume request. When we are able to reintroduce the MODs into the ODJ, we can get the database to update itself based on the contents of the MOD. When this fails, the MODs are sent to the vendor to be rewritten.

Because of database discrepancies, Scavenger was stalled, creating retry jobs on the queues for the examinations it was requesting from these partially scavenged MODs and filling up queues for system jobs. When this exceeds the limit set by the value of max_jobs, Scavenger goes back to sleep. The partially scavenged disks were taken off the jukebox and the queue cleared of retries in order for Scavenger to continue progress. This requires a more vigilant watch on AS3's ODJ queue in addition to monitoring other queues on our PACS. Discrepancies associated with the partially scavenged disks will have to be corrected by our PACS vendor before the migration is complete.

At the University of Chicago Hospitals, their PACS migration experience showed that image data in DICOM format can be faithfully propagated from a system to its successor. Their migration process was logically straightforward but primarily affected by issues of large amounts of data transferred.¹

At TCH, the migration process from one system to another also showed that image data in DICOM format may be transferred with full fidelity. Its process, though, was impeded primarily by issues related to database integrity and archive management.

The data migration experience of the Department of Radiology at the University of Utah Hospitals and Clinics, on the other hand, involved storing both DICOM and non-DICOM data in the same archive. But their experience echoes the fact that no matter how state-of-the-art the technology selected for migration, it will be made obsolete by the next generation of technology.²

CONCLUSIONS

All of us who operate totally digital radiology departments are slaves to the technology that constitutes the digital archive. Because we cannot guarantee the function of our digital archives for the full duration of our responsibility to maintain an image of record, we must be married to the concept of data migration. Migration of data is intimately associated with the image database. Because the demographic data appended to the examinations in our oldest archive was suspect, wholesale migration of those examinations was postponed until RIS validation can be performed. During the migration of our second-generation archive, discrepancies in the image database information interfered with the process. The unanticipated benefit is that data migration is forcing the total reconciliation of our image database. Otherwise, this would occur only on an incidental basis when a user could not retrieve a discrepant examination. The software tools for migrating data and for monitoring the integrity of our image database have not kept pace with the demands of our healthcare operation and need dramatic improvement.

REFERENCES

1. Behlen FM, Behlen FM, Sayre RE, et al: "Permanent records": Experience with data migration in radiology information system and picture archiving and communication system replacement. *J Digit Imaging* 13:171-174, 2000 (suppl 1)
2. Baune D, Bookman G: Migration from hierarchical storage management to ASM storage server: A case study. *J Digit Imaging* 12:75-77, 1999 (suppl 1)