Requirements for an Enterprise Digital Image Archive

Bradley J. Erickson, Kenneth R. Persons, Nicholas J. Hangiandreou, E. Meredith James, Christopher J. Hanna, and Dale G. Gehring

This report describes several image archival problems facing the authors' department and the results of their attempt to define the requirements for an enterprise digital image archive. The problems identified include the costs of supporting multiple distinct archives, the increased complexity of supporting multiple archive interfaces, the differences in data handling policies and resulting variations in data integrity, and variability in support for nonimage data. The authors also describe the data collected including image volumes and trends and imaging device trends. Finally, the resulting specification for an enterprise digital image archive, including storage and retrieval performance and interface requirements are presented.

KEY WORDS: image archive, DICOM, picture archival and communication system.

THE PAST DECADE has seen significant interest in the implementation of picture archival and communication systems (PACS) and the digital radiology department. The most visible component of PACS is the workstation. However, we believe that the archive often is an overlooked component but has great impact on the success of PACS over time. We also believe that consolidation of medical practices and increased emphasis on the use of shared services by the departments of multispecialty groups has greater implications for the archive than for other PACS components. Finally, we believe that although the workstation, database, and workflow technologies are and will continue to be rapidly changing, the archival needs are more static. By separating these, one can more easily facilitate the rapid change in the department (PACS without the A), while letting the archive maintain its stability. This report describes our recent attempts to define the requirements for an enterprise image archive, in which the first step is an implementation in radiology, with an architecture that is expandable to support other departments, including those in physically distant locations.

From the Department of Radiology, Department of Information Services, and Mayo Foundation, Rochester, MN.

Address reprint requests to Bradley J. Erickson, MD, PhD, Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN, 55905.

0897-1889/01/1402-0003$35.00/0

doi:10.1053/jdim.2001.24175

THE ENVIRONMENT—ENTERPRISE VIEW

Our institution is a large, multispecialty, multigroup enterprise with multiple facilities, with a concentration of facilities in a few locales. Approximately 800,000 radiologic examinations per year are performed at the main site. There are approximately 10 affiliated facilities with radiologic capabilities within a 100-mile radius. There are 2 larger group practices, each with "satellite" operations; each of these group practices and their satellites perform 200,000 to 300,000 examinations per year. The 2 affiliated group practices have converted to filmless operations using commercial PACS products. Their smaller surrounding practices have some digital imaging devices, and are in various stages of filmless conversion. The annual volume for the enterprise approaches 1.7 million examinations. This geographical dispersion and scale are becoming common for other healthcare enterprises.

We surveyed the imaging capabilities of our department (described below), the other radiology departments in the enterprise, and 13 nonradiology departments. The nonradiology department with the greatest digital imaging capability (in terms of volume) was cardiology.

THE ENVIRONMENT—DEPARTMENT VIEW

There were 2 main studies performed to determine departmental requirements for a digital image archive. The first was a study of how existing image archives were used, and the second was a review of plans for the implementation of new/replacement digital imaging devices.

More than a decade ago, individuals within our radiology department recognized a need for establishing a digital image archive. A joint effort between IBM and our institution resulted in an archive for computed tomography (CT) and magnetic resonance (MR) images (hereafter referred as the "CT-MR" archive) that began operation in 1991 and has been operating since that time.[1-3] In addition, the main site radiology department has implemented commercial PACS products in ultrasound (ALI

Technologies, Vancouver, CA) and in a segment of the radiography practice (GE Medical Systems, Milwaukee, WI); all systems serve in- and outpatient practices in the main site. These 3 archives store approximately 275,000 examinations per year (5.25 terabytes). Unfortunately, each of these systems has a separate archive utilizing distinct storage technologies and requiring a distinct support infrastructure.

THE PROBLEMS OF CURRENT ARCHIVES

Although these systems have been useful, they also have made us aware of several problems that have led to our desire for a single enterprise image archive. Some of these problems are: (1) Each archive is a separate product with separate support infrastructures and costs. Each archive utilizes different storage media and technologies. Economies of scale cannot be realized. (2) Each archive has a different interface, increasing the complexity of the systems that interact with the archives. (3) Each archive has different methods and abilities to validate the data being stored, which leads to variability in data quality, integrity, and, therefore, expectations. (4) Each archive has a different ability to store nonimage data.

Because the archive component of a PACS is typically integrated with a proprietary PACS database and interfaced to the PACS with proprietary software, data migration requires a special vendor-dependent application. Vendors are motivated to encourage migration to their "next-generation" PACS to generate new sales. At the same time, they are counter motivated to facilitate migration to a competitor's product. In the proposed architecture, we will remove the archive from the PACS. It will have its own internal database. The "PACS without the A" will have its proprietary database that performs workflow management and display.

Archives, by definition, must be stable and will have a long life cycle. PACS workstations, the supporting databases, proprietary communications protocols, and imaging modalities, however, tend to quickly follow the technology curve, thus having short life cycles. Ideally, the workstations and modalities could be replaced as needed with "state of the art" devices without affecting the archive. By tying the archive and workspace together, the customer must compromise between these competing needs. We recognize that achieving a pure DICOM archive may not be realistic in the near future—certain proprietary solutions will be needed to bridge the gap—but by designing the archive correctly, the transition to a fully standards compliant system will be easier.

In an attempt to mitigate these problems, our group set out to specify the requirements for an institutional digital image archive. The requirements for that archive were derived from a number of sources described below.

METHODS

Existing Archive Data

To understand archive volume requirements, the 3 existing radiology image archives at the main site were queried. For the CT-MR system, this included the period January 1, 1992 through December 31, 1999. For the ALI system, the period was January 1, 1997 through December 31, 1999, and for the GE PACS, the period was January 1, 1997 through December 31, 1999. The queries provided information on the number of devices sending to each archive and the number of examinations, images, and bytes stored on a "per-year" and "per-device" basis.

Imaging Device Data

In addition to studying existing archives, we surveyed all imaging devices in the radiology department—both those that were DICOM compliant, and those that were not. For those that did not create DICOM images, suitable replacements were identified. In some cases, the replacement will require a change in practice—for instance, complex motion tomograms of the temporo-mandibular joints are obtained daily by a device that does not seem to allow easy acceptance of a computed radiography (CR) plate. In this case, we worked with referring physicians to identify a CT procedure that would meet their needs.

We also studied trends for each class of imaging device—the number of images per examination, the number of examinations as a percentage of the practice, and total practice volumes. These data were based on both billing records and queries of the 3 archives for each year of operation. To obtain a longer perspective, we also sampled paper records for CT and MR for the years 1975, 1980, 1985, and 1990. These data were used to estimate archive storage requirements that would meet the needs of the clinical practice.

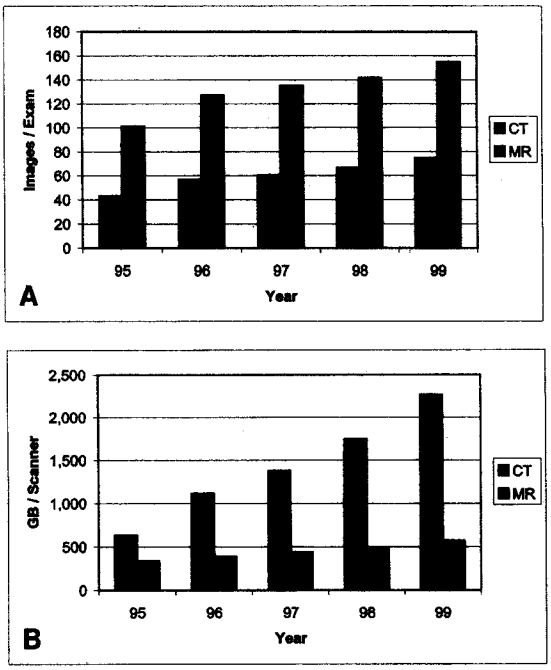

Fig 1. MR and CT examination trends. (A) The average number of images per examination per year. (B) The average number of Gigabytes of image data per scanner per year.

Retrieval Pattern Data and Image Usage Data

The retrieval requests are automatically logged for the CT-MR system. The dates and types of studies requested were retrieved from the database. In our environment, most retrievals are generated automatically by "prefetch" algorithms. It is important to determine whether the data actually are viewed because that will affect retrieval requirements. Although vendor applications generally lacked the ability to log this information, the internally developed clinical display application did start collecting this information in 1997. This display application has been described previously.[4,5] In this case, we queried the clinical display system database to determine ad hoc and prefetch retrieval volumes.

Archive Costs

The capital costs for the 3 archives were difficult to determine, because 2 were acquired as part of a PACS purchase, and the third was an internally developed archive where development costs were not tracked carefully. Therefore, the capital costs were not computed. Because archives by definition must be operated for many years, the operating costs are likely to be greater than capital costs. We calculated the operating costs (media costs, maintenance contracts, and support/operating personnel) for each of the 3 archives.

RESULTS

Main Site Radiology Departmental Needs

Imaging Device and Archive Data

After a thorough inventory, we identified a total of 147 imaging devices either present now, or that would be present in the department assuming conversion to digital (including 11 mammography devices). We are seeing an aggregate increase of about 3% per year in total examinations, with a shift from radiography to cross-sectional imaging. This means that the radiography device count is fairly static, but the CT, MR, and ultrasound (US) device count is increasing. Further, we found that the number of images per examination (Fig 1A), and the total data volume per examination (Fig 1B) have increased substantially over the past few years. This is in addition to overall increases in procedure volumes (Table 1). It is interesting to note that the increase in data volume parallels (and is driven at least in part by) the substantial improvements in computer technology. Indeed, the

Table 1. Estimated Archive Storage Requirements for Calendar Year End 1998, 2000, and 2003

	1998 Actual		2000 Estimate		2003 Estimate	
	Procedures	Data (GB)	Procedures	Data (GB)	Procedures	Data (GB)
Plain Film	407,913	1,400	400,000	7,100	425,000	12,500
CT	82,242	1,500	99,222	3,100	108,422	4,500
MRI	32,605	600	41,179	1,540	49,998	3,200
US	60,724	400	65,697	790	71,515	860
Nuclear Medicine	57,902	50	61,628	280	64,023	290
Angio/Fluoro	51,264	0	51,881	1,300	55,934	1,400
Mammography	46,128	0	49,887	0	54,513	11,000
Total	738,778	3,950	769,494	14,110	829,405	33,750

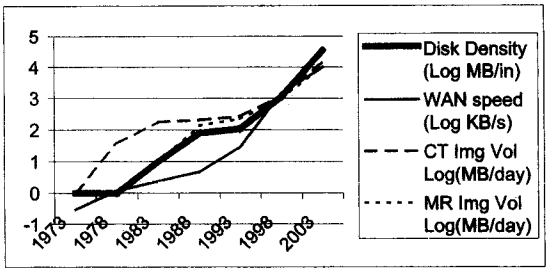

Fig 2. Comparison of image volume generated on a per-day basis over time with network and disk storage capacity increases caused by technology improvements. Over a period of many years, these tend to parallel each other, reflecting the fact that the technology driving the advances in the computer industry also drive advances in imaging technology.

curves parallel storage and network capacities over large time intervals (Fig 2).

Retrieval and Usage Patterns

We found that the archives were retrieving a large number of examinations. For instance, data from the archives showed that about 165,000 retrievals were performed in 1999, versus the 275,000 new examinations collected in that period. Figure 3 shows the distribution of the age of examinations retrieved over this period. These data may be skewed because only the CT-MR archive had depth substantially greater than 2 years.

One significant challenge to assessing the value of this information is that many retrievals are generated automatically. An important question is how many of the images retrieved actually are reviewed. We were able to estimate this because the clinical display application provided a mechanism to record whether examinations actually were viewed. On an average day, 148 examinations are retrieved automatically to the clinical image viewing system to provide a historical comparison examination (prefetched), whereas 20.2 examinations are requested for "ad hoc" retrieval. The difference between the number of retrievals quoted here (168.2) and the total retrievals quoted above (165,000/year = 452/day) is caused by other systems, such as the radiology image review systems and research. The departmental systems tend to generate more retrieval requests because they have very limited online storage capacity. The clinical viewing system has an online capacity for about 6 months of new patient examinations. Overall, only a very small fraction (less than 10%) of prefetched images currently are viewed electronically by referring clinicians. For reference, however, this is

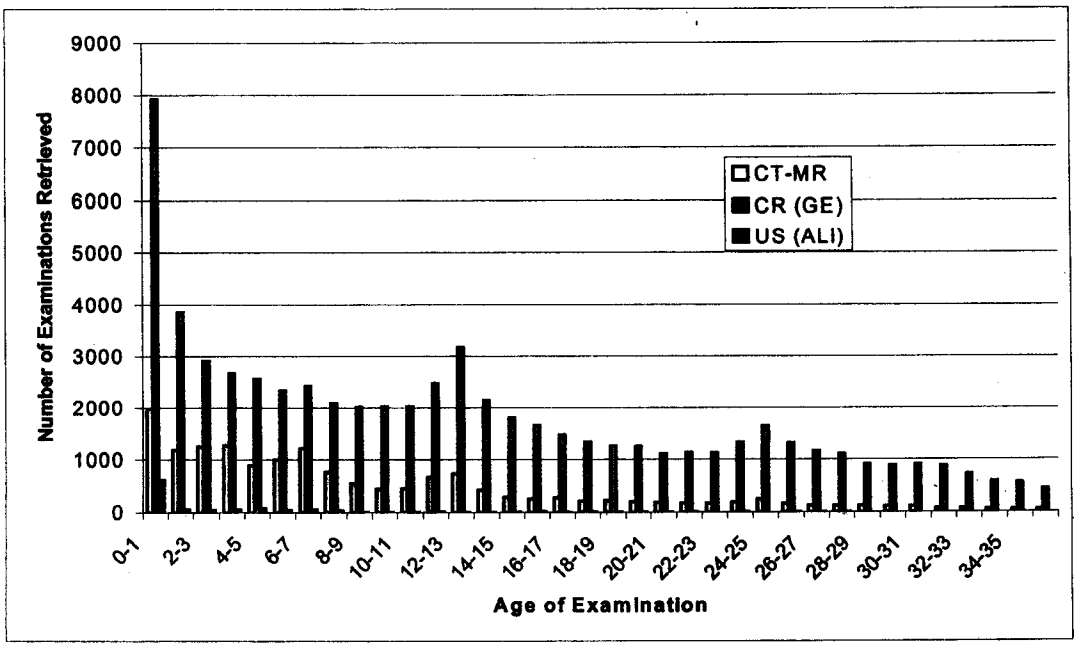

Fig 3. Age of examination versus query volume. One can see peaks at 1 and 2 years, which correspond with clinical practice of 1- and 2-year follow-ups. The CR practice performs a higher proportion of inpatient examinations than the other practice areas.

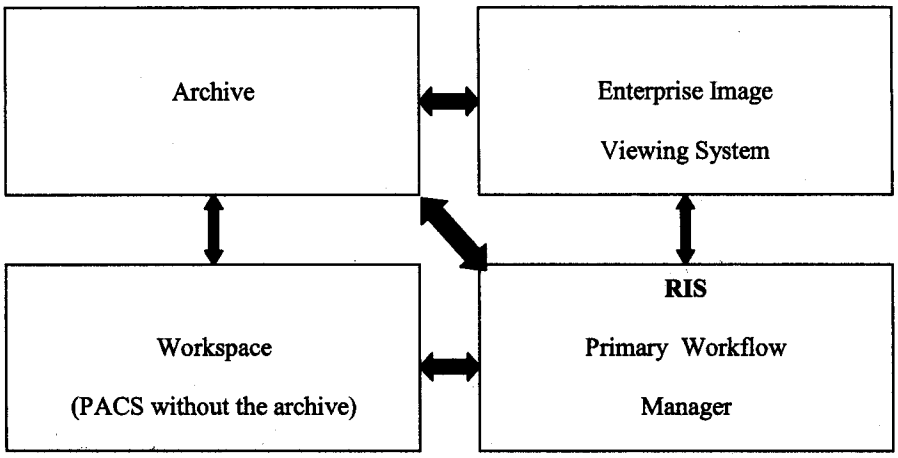

Fig 4. Schematic diagram of the system architecture and data flow with a centralized DICOM archive. Note that the archive only responds to requests for data by other systems that control workflow. In this way, the differing workflow needs of other departments, or upgrades to existing departmental systems can be added without requiring a change to the archive. DICOM image data move between the archive and the workspace (PACS without the A). Examination information and move requests are passed between the RIS and the archive. Examination status and worklists pass between the RIS and workspace. The image viewing system receives images from the archive, and examination information from the RIS.

not substantially different from the rate of viewing of new examinations in our environment.

While the archive must service the transfer volumes noted above, we feel that by design, the archive should not be controlling the workflow. That means that prefetches and worklist generation will not occur within the archive. Figure 4 shows the data flow in the proposed system architecture. Further, we (and others) have found it difficult to implement successful prefetching algorithms. Therefore, based on the pattern we saw, where most images viewed were less than 2 years old, we plan to have that on-line capacity, and never implement prefetching.

Archive Operational Costs

The annual operating cost for all 3 archives was estimated to be $943,199 in calendar year 1999. This consisted of 3 components: media, maintenance contracts, and support and maintenance personnel.

All 3 archives utilized optical disks—one used 14-inch media, whereas the other 2 used 5.25-inch media. The density of the media also varied. The cost of raw storage (ie, not considering variations caused by image compression), was very similar—4.4, 4.0, and 3.0 cents per megabyte. The combined media cost for the 3 systems was $212,949 in 1999. (Although tape systems were not available for any of these PACS at the time of purchase, their raw media costs are much less—typically in the range of 0.5 cents per megabyte—although the lifetime probably is less, which might increase true cost. Of course, newer optical disk storage technologies also are much less expensive.)

Maintenance contract costs were calculated directly from vendor maintenance agreements for the hardware device (not the software). All 3 were similar, which is not surprising given that all 3 used optical jukebox technologies. The sum of these was $245,000 for calendar year 1999.

The personnel costs were the most variable. This is because the implementation (eg, off-line storage use) and age of the technologies was substantially different between systems. One system supported off-line media, and consumed 0.3 FTE (full-time equivalent) of a computer operator just for media handling. Engineering and programming support for the 3 systems also varied, totaling 6.5 FTEs. When the fully burdened cost for each category of FTE is factored in, the total personnel costs were $485,250 in calendar year 1999.

Enterprise Requirements (High Level)

As noted at the outset, the purpose of this work was to develop a requirements specification that could be used to develop a design that could meet the needs for a large radiology department in one city, and which could be expanded to accept im-

ages from other departments and from other locations in the enterprise. An important finding was the "archive data volume" from departments likely to be producing a substantial number of digital images in the next 5 years (both in the main location, as well as the group practices and "satellites"). The 5-year cumulative data storage requirements would be approximately 800 terabytes if all digital images (static and dynamic) were to be stored in uncompressed format. This did take into account growth trends in the number of examinations performed per year but did not assume conversion of current analog imaging devices to digital beyond what was currently planned, nor did it estimate the increased number of images in each examination when digital imaging becomes available (the so-called "digital diarrhea" effect). This requires an architecture that can scale up well into the petabyte (2^{50} bytes) range.

One important advantage of the multisite capability is the ability to mirror data sets, which can be a valuable protection against disasters. In evaluating this, we also found that different facilities had different expectations for recovering from disasters. We also found that "recovery" to an operational condition was not a binary event (eg, hours to become 'minimally' operational, and hours to become fully operational) requiring a more complex specification of disaster recovery.

Resulting Functional Requirements Specification

As a result of this research, and the need for an institutional digital image archive, the functional requirements for the archive were developed. The first phase of implementation would satisfy the needs of the main site radiology department over the next 5 years. The functional requirements include the following important points.

1. Storage Volume

Based on the data volumes projected in Tables 1 and 2, the expected capacity for the system must be 5 million examinations of rapidly accessible storage, constituting 96 terabytes (TB) (before any compression). Note that this excludes mammography until the year 2003, when we assumed a phase in with fully digital acquisition by end of 2004. A conversion of all other currently nondigital imaging devices is assumed to occur during calendar years 2001 and 2002. Image volumes outside the department from devices that either were digital or were expected to be digital in the near future suggested volumes in the 100-200 TB per year range. Adding in the facilities in the region added another 20 TB per year currently, with a substantial volume of analog imaging continuing for the foreseeable future.

2. Archive Performance

The archive must be able to handle the daily storage volume without impeding the daily operation of the imaging modalities. The radiology department should be free to archive these data on an ad hoc basis (as the images are generated) or as a scheduled batch process (eg, overnight). The archive must be able to be configured to support at least 300 imaging devices, with up to 100 simultaneous archive associations active at a given time, regardless of the number of retrieve associations

Table 2. Archive Transactions and Data Per Hour

	1998 Actual		2000 Estimate		2003 Estimate	
Modality	Transactions/h	Data (GB)/h	Transactions/h	Data (GB)/h	Transactions/h	Data (GB)/h
Plain Film	185	1.33	207	7.45	212	7.63
CT	45	6.75	51	11.47	54	12.15
MRI	15	2.25	21	3.75	22	3.30
US	24	0.48	34	0.68	36	0.72
Nuclear Medicine	21	0.04	31	0.06	32	0.062
Angio/Fluoro	0	0	27	0.27	28	0.28
Mammography	0	0	0	0	27	1.94
Total	290	10.85	371	23.68	411	26.082

NOTE. *Exam Assumptions:* Data equivalents for CR = 9 MB per image, 2 images per examination; CT = 0.5 MB per image, 150 images per examination; MRI = 0.25 MB per image, 150 images per examination; US = 0.5 MB per image, 40 images per examination; NM = 0.2 MB per image, 10 images per examination; Angio/Fluoro = 1.0 MB per image, 10 images per examination; All other images = 0.5 MB per image, 2 images per examination.

that are active. Table 2 shows an estimate of the number of archive transactions and data volumes per hour by modality type for the years 1998, 2000, and 2003.

3. Storage Life

The main site has statutory requirements to keep medical images for a minimum of 5 years if normal and 7 years if there is a positive finding. Exceptions to this rule are mammography (indefinite), pediatric (until the patient reaches age of majority), and obstetric (age of majority) diagnostic images. Our institution's philosophy is that the image data are a primary corporate asset, and, therefore, all images should kept for substantially longer than 7 years, and some data should be kept indefinitely to maximize education and research opportunities.

4. Archive Storage Interface

The archive must conform to the DICOM standard in areas relating to storage of objects (images, overlays). It should accept DICOM storage (C-Store) requests as "service class provider" (SCP) for all radiology-related service-object pair (SOP) storage classes (phase one) and all SOP storage classes (phase two). The archive must support both the "push" and "pull" models of DICOM storage commitment as service class provider. The pull model of storage commitment also requires that the archive support DICOM query/retrieve as "service class user" (to retrieve the images to be stored from the image source). Figure 4 shows the data flow and architecture for the radiology department implementation.

A DICOM conformance requirements document was developed to further specify the details of the DICOM archive interface requirements. Although DICOM meets many of the interface needs, there are other (HL-7) interfaces required to handle patient demographics. It appears that the "Integrating the Healthcare Enterprise" or IHE initiative may address some of these needs (see *http://www.rsna.org/IHE/ihe_index.html*).

5. Retrieval Volume

The archive must be capable of accepting the volume of DICOM C-Move transactions per hour specified in Table 3, and meet the retrieval performance requirements described below. The archive must be able to support at least 300 retrieve devices, with up to 100 simultaneous retrieve associations active at a given time, regardless of the number of archive associations that are active.

6. Retrieval Performance

Retrieval performance requirements for radiology are specified here for ad hoc retrievals of images that are less than 2 years old and images that are more than 2 years old. These retrieval performance requirements specify the minimum performance requirements for the system under peak load conditions (see retrieval requirements in Table 3).

For example, during the busiest hour, when the system is servicing an estimated 1,229 retrieval transactions per hour (Table 3), a user should be able to perform an ad hoc retrieval of an examination in a reasonable time. The "reasonable" re-

Table 3. Ad Hoc and Prefetch Retrieval Transactions Per Hour

	2000 Estimate			2003 Estimate		
Modality	Prefetch/h	Ad Hoc/h	Max/h*	Prefetch/h	Ad Hoc/h	Max/h*
Plain Film	414	124	663	425	127	680
CT	0	0	0	108	33	173
MRI	0	0	0	45	13	72
US	67	20	108	72	21	114
Nuclear Medicine	62	19	100	64	19	102
Angio/Fluoro	0	0	0	55	16	88
Mammography	0	0	0	0	0	0
Total	543	163	871	769	229	1229

*NOTE. The following assumptions were made for the retrieval volume calculations shown in the table above: average study size, 30 MB; Max hour, Assume 40% of ad hoc retrievals occur during 1 busy hour, ½ of radiology department retrievals prefetched in a 4-hour period overnight, ½ of radiology department retrievals done ad hoc randomly over an 9-hour day with 60% of those done over an 8-hour period, and 40% done in 1 busy hour; CT, MRI, angiography and mammography studies will not be retrieved in 2000 because there will be no installations of PACS for these modalities in this period.

trieval time (time from issue the DICOM request to begin of the transmission of image data) is:

Once a retrieval is begun, the archive should be able to sustain the image transfer at a mean rate of no less than 5 MB/sec, assuming adequate network bandwidth and receiver capacity.

7. Query/Retrieval Interface

The archive must conform to the DICOM standard in areas relating to the query and retrieval of stored objects (images, overlays). The archive should support (as service class provider) a direct DICOM query (C-Find) at the patient or study level, and a retrieve (C-Move or C-Get) at the study or series level using either the study or series unique identifier (UID) as a unique key. All images for that series or study should be retrieved. The archive should support the priority attribute on the C-Move and C-Get requests. The archive should implement a retrieval queuing mechanism based on low-, medium-, and high-priority requests. For example, an information system (such as an RIS), which knew the specific study or series UID that it wanted to be retrieved, could issue a DICOM C-Move at medium priority to the archive specifying that study or series UID, and the application entity title of the destination device where the images should be moved. The archive should honor that request ahead of any low-priority retrieval requests that are not already in process.

The archive must support both the push and pull models of DICOM storage commitment as service class provider.

8. Other Requirements

(a) Data compression. Multiple copies of all images will be written to low cost media (such as tape) that can be stored in a physically remote site, using reversible compression. To reduce storage costs while achieving high performance, the archive also must be configurable to apply compression (reversible or irreversible) at an appropriate compression level for a given type of image. We expect that the routinely accessed images less than 2 years old will be stored in irreversibly compressed format on magnetic disk. The archive should be able to receive and send DICOM objects using the DICOM-defined transfer syntax options that incorporate compression. A special interface will be provided to allow access to losslessly compressed images for research, education, and medico-legal purposes.

Images older than 2 years (capacity of the irreversibly compressed component) will be retrieved from the reversible format and compressed using the most current compression algorithm. In this manner, we are able to "roll over" to new compression schemes and need only support a 2-year window of irreversible compression technologies.

(b) Legacy data. Over 50 million images exist in the current CT-MR archive system. These legacy data will be migrated to the new archive in DICOM 3.0 format. All images acquired in non-DICOM format (eg, the CT-MR archive has many images in ACR-NEMA 2.0 format) will be converted on exportation by the source archive.

(c) Data access. Only systems that have institutional approval should be able to store images, query the archive, or gain access to archived data. The archive should maintain a list of trusted AE partners and their permissions. Only authorized system administrators will be able to access the system via a system provided user interface. The archive data should be accessible in a hierarchical fashion using the following categories: on-line (RAID), near on-line (in a library device not requiring human intervention for access), and off-line (not in a library, potentially in a physically distant location).

(d) Data confidentiality. As with other aspects of the patient record, patient confidentiality of archived data must be protected. The systems that have access to the archive will have security in place. Users that have permission to sign onto those "trusted" systems also will have access to the archive resource as their system permits, whereas those who do not have appropriate permissions will be refused.

(e) Archive reliability. To the greatest extent practical, there should not be any single point of failure in the archive system. Only in the case of a multibuilding disaster should the archive be inoperable. Multiple, time-stamped copies of the data should be made to protect against errant software, and these copies should be stored in physically remote sites for recovery from large-scale catastrophes.

Another aspect of reliability is guaranteeing that what goes in is what comes out. We therefore require that each data set be "retrieved" immediately after storage, and a bit-by-bit comparison be

made to ensure that the image can be accurately retrieved (and decompressed if it was compressed).

(f) Data availability. The data must be available when and where it is needed. Lack of availability of the data would affect the institution's ability to care for the patient, and would be considered a serious problem, even for short outages. The solution that is put in place should be a "continuously availability solution" with redundant copies of the image data online, and automatic fail-over hardware/software components in place.

The archive must provide archival services within 24 hours of a catastrophic (ie, multibuilding) failure. It must provide query and retrieve services for 2 years of images (in irreversibly compressed format) within 24 hours. The rest of the image database must be available within 72 hours.

(g) Correct information in exported DICOM image headers. Images exported from the archive should contain the correct accession number for that examination, and the correct demographic information based on the most up-to-date known demographic information for that patient. When corrections are made (demographic, accession number) to images that are stored in the archive, those corrections should be included in the DICOM header when the image(s) are exported. This will be implemented using an HL-7 interface to the enterprise registration system. In the event that the registration system is unavailable, the archive should service the request immediately using the last known good values.

(h) Workflow.

- RIS Interface With the Archive

Among its other responsibilities, the main site RIS provides for the "management" of the radiology information stored in the archive. Today, it orchestrates prefetch retrievals from the archive in preparation for examinations for which it has an order. To do this effectively, the RIS needs to know what is stored in the archive (what studies, RT plans, standalone overlays) for the historical examinations it knows about. In our desired DICOM environment, the RIS would know what instances (images) were created by the modalities through the DICOM-performed procedure step. The RIS would use DICOM storage commitment to know which instances were stored successfully in the archive. Until these DICOM interfaces are in place (in all of the modalities, and in the RIS) an interim solution is for the archive to provide a private interface to provide the RIS with DICOM series and study UIDs (unique identifiers) each time it stores an instance (study, RT plan, overlay).

We are hopeful, however, that most, if not all, prefetching will be eliminated by having 2 years of information on-line. It is possible that cross-site sharing, and some well-defined clinical scenarios will drive the need for limited prefetching, however.

Other private interfaces (using HL-7) will be needed to insure changes made to the examinations and patient information (demographic changes and corrections, deletions, merges, splits) are known by both the RIS and the archive.

- Modality and PACS Integration With the Archive

We use the term *PACS* here to include quality control workstations, interpretation workstations, and the associated database and RAID that ties them together (eg, PACS without the A). Image data, such as graphical information, annotations, and key images, must be able to be archived and exported in DICOM format. This requires that PACS provide this information in DICOM format.

(i) Data migration from aging media. To enable the system to take advantage of future technologies, the archive provider must have a strategy for dealing with migration of data to new media and storage devices without losing access to the data during the migration process. Migration of the database also must be considered. The database should be built with the ability to export all non-image data using a self-describing ASCII-text format (such as XML), increasing the likelihood of successful export of data to a future database technology.

(j) Rearchiving. If the same image is archived more than once (eg, the image has the same instance UID as an image that is already stored in the archive) the original copy of the image should be maintained by the archive, and the more recently received copy should be discarded. *If a series is rearchived and it contains the same images plus additional images, or if new images come into the archive for a series that is already archived,* the archive should add the additional images (not overlay the duplicate images) and inform the RIS of the new image count for that series via the same

interface it would use to report a newly archived series.

9. Administration

The system should provide an administrative tool to move, copy, and delete images, and to delete and create series and examinations. This should be a graphical tool where images can be viewed and can be dragged graphically into and out of examinations or series folders. The ability to alter accession numbers and associated patient ID (internal ID) will be needed. Changes that affect RIS examination management need to be communicated to the RIS. The archive should maintain an audit trail of these changes.

(a) Archive verification reports. The archive must provide an archive verification report via a web-based application for imaging devices that do not support DICOM storage commitment functions. The archive verification report should enable the technologist to determine if an examination (or any part of an examination) has been archived successfully, so they can safely delete the examination from the imaging device or workstation.

(b) Diagnostic and administrative reports. The archive must be able to generate administrative reports including archival (number of images, series and examinations, total digital data stored—by modality type, by device, by time range, and by procedure) and retrieval (number of images, series, and examinations retrieved, age of examinations being retrieved, retrieval performance—by modality type, by user, by target device, and by time range). For diagnostic purposes, the archive should provide automatic system monitoring for performance, as well as transaction and service logs.

DISCUSSION

Although filmless operations in radiology remain relatively uncommon in the United States, an increasing number of institutions have adopted electronic imaging, at least within the radiology department. Unfortunately, a review of the commercially available PACS products found a significant focus on the user interface of the diagnostic workstation and relative ignorance of the archiving device. Some implementations have suffered severely from poor implementations of archive devices—most PACS archive devices were designed as "backend integrations" using off-the-shelf storage devices designed for infrequent retrieval. Anyone familiar with the practice of radiology knows that old examinations are valuable, as evidenced by the common saying "the historical film is the radiologist's best friend." This also is seen in practice by the high rate of image retrieval even in our film-based institution.

Commercial products continue to focus on departmental needs (ie, how good is the diagnostic workstation) without the vision for a single institutional archive. This implies that each department will have its own solution, develop its own interfaces to share information, and manage later migration to new storage and database technologies.

The value of providing an institutional image archive as a separate entity has been described previously by Erickson and Hangiandreou.[6] Such a design minimizes the number of interfaces that need to be developed by the institution has a whole. It also provides a more unified view to the clinical user. Finally, and perhaps most importantly, the infrastructure requirements to maintain and update the archive device can be managed more efficiently, particularly if workflow components are not incorporated in the archive.

It may be difficult to justify an electronic image archive if there is no electronic image distribution system. However, the authors felt that it was logical to first implement the archive, and as it was loading, implement soft-copy interpretation. In this way, when the workstations are in place, filmless interpretation (both new and compare examinations are electronic) can be immediately implemented without requiring film digitization or film handling, because there will be 1 to 2 years of digital examinations ready for comparison purposes. This delay also is necessary because we have not found a current PACS product that satisfies our clinical practice, although we are working with a vendor, and anticipate a solution in that time frame. It is possible that we will have a partial filmless operation, where we eliminate CT, MR, and US films before going completely filmless. Although this helps avoid the problems associated with film digitization, it places greater demands on the verification and validation of data, because the images being archived are not checked visually by any person after being sent by the source imaging device. However, few, if any, archives check data after the viewing station, and so this is not unusually risky. We also believe that some advantage

will be gained during this period if a clinical image distribution system is connected to the archive.

Although a major component of this document describes seemingly mundane collections of data like examination volumes, imaging device counts, and image viewing rates, as well as initial arbitrary requirements for performance, these are the criteria by which success for an archive will be determined. Failure to agree on the specifications and how they will be met will result in failure to protect the imaging department's greatest asset—the images and associated information that they produce. It raises the interesting question of how one should measure archive quality. We will measure both the simple ability to handle the data archival and retrieval volumes for the many different types of images, as well as less binary properties such as fault tolerance, data verification methods, and suitability of interfaces.

The use of irreversible compression for medical images is controversial. The primary concern is medico-legal exposure. In the design specified, the "original" digital data will be retrievable from a tape storage system for the rare case in which there is concern that irreversible compression may have "altered" the appearance or might affect research outcomes. However, routine clinical work (eg, distribution to clinicians and historical comparisons) will use irreversibly compressed image data stored on magnetic disks. We believe this design mitigates the potential legal exposure while substantially reducing cost. We estimate that the savings for our institution by using conservative irreversible compression will be more than $10,000,000 over the next 5 years. Some argue that the advance of technology will reduce the cost, mitigating the need for compression. However, Fig 2 clearly shows that the same advances in storage and network technology on which these arguments are based, also drive the imaging devices to create larger data sets.

CONCLUSIONS

For decades, imaging in the medical community has been done with film. To manage that film, hospitals and departments traditionally have created filing systems, storage areas, and methods to transport and track the distribution of film to and from locations where it is needed. With the rapid spread of digital imaging in medicine (digital cameras, digital radiology systems, and digital catheterization laboratory systems) comes corresponding need to manage and store the terabytes of digital data that are piling up on departmental imaging systems. Paralleling this migration from film-based to digital-based imaging systems is the development of enterprise systems to record and distribute other information (notes, reports, lab results) to the physicians desktop.

We believe that both of these initiatives (migration from film to digital, and the creation of electronic medical record distribution systems) are best served by a centralized DICOM archive to store and manage graphic information such as images overlays, radiation therapy plans, and annotations. Medical images are the hard assets (eg, nonpeople) of an imaging department. Failure to architect a solution that thoroughly protects this asset while making the information readily available to those that need it puts the radiology department and the institution at risk. Providing this as a single, centralized service reduces cost and complexity.

REFERENCES

1. Hangiandreou N, Williamson B, Forbes G, et al: Clinical implementation of the Mayo-IBM PAC system. Radiology 189:290, 1993

2. Hangiandreou N, Erickson B, Williamson B, et al: Current status and future directions of the Mayo-IBM archive system. Radiology 197:257, 1995

3. Gehring D, Persons K, Rothman M, et al: Detailed description of the Mayo/IBM PACS SPIE, Newport Beach, CA, 1991

4. Erickson BJ, Ryan WJ, Gehring DG, et al: Image display for clinicians on medical record workstations. J Digit Imaging 10:38-40, 1997

5. Erickson BJ, Ryan WJ, Gehring DG, et al: Clinician Image Review Patterns In An Out-Patient Setting. SPIE Medical Imaging, San Diego, CA, Feb 24-28, 1998

6. Erickson B, Hangiandreou N: The evolution of electronic imaging in the medical environment. J Digit Imaging 11:71-74, 1998