PACS Databases and Enrichment of the Folder Manager Concept

Katherine P. Andriole, David E. Avrin, Lloyd Yin, Robert G. Gould, and Ronald L. Arenson

Current challenges facing picture archiving and communication systems (PACS) center around database design and functionality. Workflow issues and folder manager concepts such as autorouting, prefetching, hanging protocols, and hierarchical storage management are driven by a properly designed database that ultimately directly impacts the clinical utility of a PACS. The key issues in PACS database design that enable radiologist-friendly, cost-effective, and data-secure systems will be discussed, including database difficulties of the DICOM standard, HIS/RIS/PACS (hospital information system/radiology information system) connectivity, and database issues in data acquisition, data dissemination, and data display.

KEY WORDS: picture archiving and communication system, database, folder manager, workflow manager, hospital information system/radiology information system.

DICOM has proven to be a remarkably successful standard for moving image and associated patient data from imaging modalities to a (picture archiving and communication system) PACS. However, for the core components of a PACS and integration with workstations, DICOM alone is not sufficient. Most if not all successful scalable PACS rely on relational database technology, often in a client-server configuration.

This report explores issues related to the relational database component of a PACS that have arisen at our institution in both our in-house constructed system and a commercial system. Folder manager concepts such as autorouting, prefetching, hanging protocols, and hierarchical storage management, developed by Arenson et al,[1-4] are in fact even more critical today in realizing the full functionality and clinical utility of these systems. This topic has not been re-addressed in the literature since the advent of relational databases and DICOM. Important issues in PACS database design that enable radiologist-friendly, cost-effective, and data-secure systems will be identified.

Figure 1 shows the overall data flow and corresponding data protocols used from an examination scheduling event through patient arrival and image scanning, distribution, interpretation, and archival. Once the patient arrives at the hospital registration area, an event is sent to the HIS. Upon scheduling a diagnostic imaging procedure, an event is sent to the RIS, and patient demographics, medical record number (MRN), patient location (etc) are sent via HL7 (Health Level or Language 7) protocol from the HIS to the RIS.

On the day of examination, patient demographics are sent from the RIS to the imaging device (ie, computed tomography [CT] scanner) via the DICOM Modality Worklist feature. Once the study images are generated at the modality, they are sent to the PACS acquisition gateway device using the DICOM communication protocol. An HL7 message is sent from the PACS to the RIS to verify the study demographic data including MRN, accession number (AccNum) and examination type or mnemonic (ExamMNE), and then passed back via HL7 to the PACS. The examination is inserted into the PACS database, typically using SQL (system query language), routed to a display station via DICOM, and sent to the archive using DICOM.

After the radiologist has viewed the image at the workstation and interpreted the examination, the report is generated in the RIS from HL7 messages originating at the dictation system. (In time, this procedure could change to direct transcription in the RIS using voice recognition at dictation.) The RIS can then communicate the report to both the PACS and the HIS using HL7 protocols.

DATABASE PHILOSOPHICAL ISSUES

Database Difficulties of DICOM

The DICOM data model is inherently image-centric, and includes patient demographic and examination level data as part of the image header, but DICOM does not specify a relational or other database schema using these parameters. Therefore, patient and study (or examination) level queries are difficult in DICOM, and are not a truly intrinsic portion of the data model. Because radiol-

---

From the Department of Radiology, Laboratory for Radiological Informatics, University of California, San Francisco, CA.

Address reprint requests to Katherine P. Andriole, PhD, University of California, Department of Radiology, Box 0628, 530 Parnassus Ave, San Francisco, CA 94143-0628.

0897-1889/00/1301-0002$10.00/0

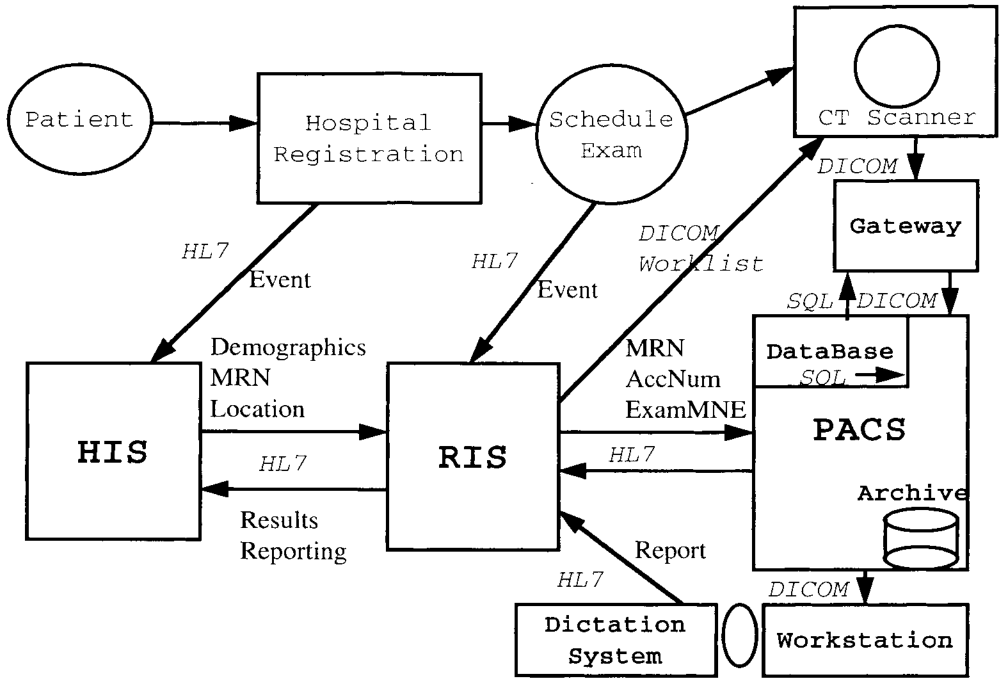


Fig 1. Data flow and intercommunications between the component systems.

ogy in the practice of medicine is patient- and study-focused, this causes a problem. The most successful solution to this problem is the addition to the DICOM application entities of a relational or object-oriented database that is patient- and study-centric, and that also contains the finer granularity information associated with image series or sequences, as well as the individual images. A typical relational database schema is shown in Fig 2.

This schema is patient-centric through the pa-

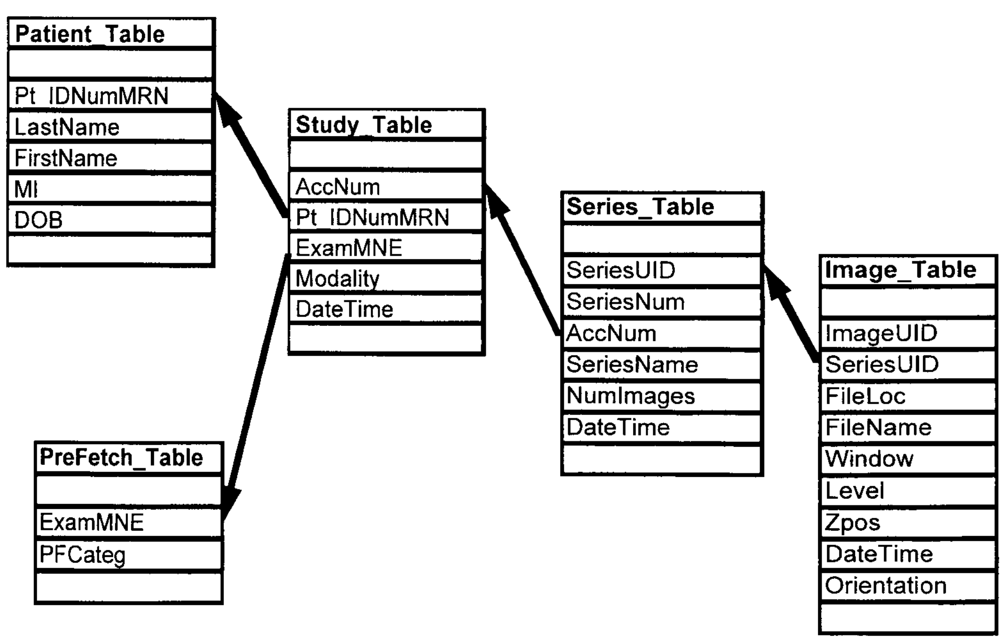


Fig 2. Patient/study-centric relational database schema for PACS.

tient table (Patient_Table). It is also examination-centric, through the study table (Study_Table), which relates each examination record to a particular patient in Patient_Table, through the patient identification or medical record number (Pt_ID-NumMRN). Each examination can consist of 1 or more series of images (a series can consist of 1 or more images), and this detail is contained in the series table (Series_Table). Series information is linked to the study through the examination accession number (AccNum).

In the formal DICOM model, every image (even for a multi-series, multi-hundred-image magnetic resonance [MR] study) is a distinct object, and in practice a distinct file. So at a minimum, the fine granularity image table (Image_Table) of a PACS within a relational database, contains the file names of each and every individual image. This table can become populated with a huge number of records. (Consider for example the sum of all the images, in each of the series, for each patient having had an MR or CT scan over the past 5 to 10 years. For even a medium-sized community hospital, the number of table entries could be 25 to 100 million!)

A very large number of database interactions (queries and retrievals) is required just to acquire and move a single study. Note, however, that very little of the header information changes from image to image within a series. For example, items such as patient name, scanner model, and institution name remain the same for the entire examination, whereas parameters such as z-axis position, time of acquisition, and possibly window and level may change from series slice-to-slice (or image-to-image). Thus, while DICOM has provided a viable image standard format and a communication protocol language for getting image exams *into a PACS,* there are database concepts and requirements outside the DICOM standard that, if implemented, could provide a more efficient and effective means for moving images and data *within a PACS.*

HIS/RIS/PACS Connectivity

One way to resolve the mismatch of the DICOM image-centric philosophy with the patient- and study-centric philosophy of clinical operations is through HIS/RIS/PACS connectivity. Health care institutions' patients each have a unique identity, such as a patient identification number (ID) or medical record number (MRN). Patients have examinations, and each one must be uniquely identified, such as by an examination accession number (AccNum). The examinations contain 1 or more series, and each series contains 1 or more images. These can be uniquely identified by a series and image number (SeriesUID and ImageUID), respectively. Interfacing the study-centric HIS/RIS database with the image-centric PACS can provide an overall patient-centric database in which to function clinically. This is the core database schema of most successful clinically functioning PACS and is shown in Fig 2.

The first requirement of the interface between a PACS and an HIS and/or RIS is to uniquely and correctly identify every patient. A troublesome consequence of ambiguous patient identification is inability to retrieve previous studies on a current patient. Unique identification is almost always accomplished via an MRN rather than by social security number (SSN). (Newborns, immigrants, and patients of lower socioeconomic status may not have an SSN.) Use of a name can be fraught with spelling inconsistencies such as middle names being written out versus using middle initials only.

The MRN usually is stored in text rather than integer format, and must be unique and sortable. When interfacing to other databases, the MRN can be miscommunicated if one system represents the MRN in integer format and the other in text format. This can create the "leading zeroes problem" mismatch. For example, the MRN represented in text format as 0012345 is represented as 12345 in integer format. One solution to the mismatched representation is to implement the capability for the integer-representation database to add back the leading zeroes to the MRN when communicating with the text-representation database, which must have an exact text representation match. For the reasons mentioned above, it is preferable to represent the MRN in text format.

Significant events in the HIS world include "patient admitted" and "patient discharged," as well as changes in patient location. HL7 has the reputation of being a weak standard, and some say not a standard at all. However, it does provide a text-based exchange between different health care databases. A new and reportedly much improved version will be released soon.[5] RIS connectivity provides notification of key events in the radiology world such as examination scheduled (canceled or changed), completed, dictated, etc. The most successful clinically implemented PACS are tightly interfaced to and integrated with a health care HIS and/or RIS.

Database Schema

The four principal tables (patient, study or examination, sequence or series, and image) are illustrated in Fig 2. When implementing such a schema in a PACS, two issues are critical. First, the unique and correct accession or examination number is the single most important piece of data within radiology. It uniquely identifies a study and the series contained in that study. It represents the fundamental unit of work or transaction within the radiology department. It usually has a one-to-one correspondence with the examination report. Second, the series table requires the creation of a (sequential) index or key to assign to each series (SeriesUID) so that the series are uniquely identified. Image level data has an unique identifier (ImageUID) assigned by the imaging modality.

As previously mentioned, a scalability problem currently being encountered in PACS databases is the size of the image level table (Image_Table) and the number of transactions required to acquire or move a study. As MR and CT examinations continue to move in the direction of more images per study, this problem becomes more severe. Note, however, that the records in this table for a given series are largely redundant, with the exception of predictable changes in the file name (eg, ex2253series02image001, ex2253series02image002, etc.), time of acquisition, and *z*- (or *x*- or *y*-) axis position, which also is contained in the DICOM header. Other image-specific data such as window and level are more problematic because they may be optimized for viewing by a technologist, radiologist, or algorithm, with the desire that they be saved for subsequent viewing, yet are not "writable" to the DICOM header. This is variable for MR, in which window and level may change from image to image within the same series. Other than the fact that series are of variable length, and can be very large, it may be worthwhile to store in a fixed-length array in the series table, the window and level values for each image.

The image table scalability issue may become manageable by determining what is really needed from the header, which header elements need to be updated, and which header elements are redundant. Note, however, that this may then create problems when adding selected images, virtual summary series, or overlays to an examination. Except for query and movement of image data, it may not be possible to avoid the image table. Transactions with the Image_Table may be reduced, however, by a filename wildcard search (ie, filename.*). Another approach, from a previous non-DICOM database design in our laboratory, used concatenated series files of images and headers, with the headers extracted each time the series was used.

DATABASE ISSUES IN DATA ACQUISITION

Bidirectional Monitoring of Acquisition

The well-designed PACS holds newly acquired studies in a restricted area (fix-queue or "penalty box") until the demographic data in the header is matched to a pending examination request from the RIS/HIS. If any failure occurs, such as an incorrect MRN or DOB (date of birth), the new examination will not pass automatically into the system to be archived (although it may be displayable) until the discrepancy has been resolved by human intervention. However, the inverse test has not been implemented. Pending examination orders held in the RIS/HIS that do not relate to any incoming PACS image data within a certain time frame should be flagged. Full PACS acquisition quality assurance (QA) requires this bidirectional process monitor to ensure that data in the PACS are valid and verified with data in the RIS/HIS, and that all data in the RIS/HIS are acquired into the PACS. This also may assist QA procedures in determining which studies have been ordered and completed but have no associated report, and therefore may not have been read. Examinations that have been canceled or performed as a different examination than originally ordered must be carefully updated in the RIS, or they will persist unmatched by an imaging study.

PACS Versus RIS Database

During the transition from a film-based to PACS environment, there is a frequently recurring problem of the PACS being ignorant of previous film-based studies on a given patient because the PACS database is typically a subset of the RIS database, containing only those studies in the RIS that begin with the existence of the PACS. In addition, RIS interfaces do not currently clearly distinguish a film-based examination from a PACS-based examination.

Operationally, this compels the user (radiologist or fileroom technician) to check not only the PACS examination history, but to independently and additionally check the RIS for non-PACS relevant previous tests. This can impinge on personnel efficiency gains in the PACS cost-justification model (and in reality!), and compromise the efficacy and economics of a PACS installation. The only solution is tighter HIS/RIS/PACS integration so that ultimately they function as virtually the same database.

DATABASE ISSUES IN DATA DISSEMINATION

Folder Manager: Autorouting, Prefetching, and the Metatable

Many of the software intelligence issues for PACS derived in an earlier era of slow networks still are relevant today, and can be addressed by the Folder Manager Concepts (1-3) that include autorouting, prefetching, and hanging protocols. Autorouting and prefetching often are mentioned together, although in fact they are unrelated. Autorouting is the ability to send an examination to the correct workstation(s) for interpretation and review. Prefetching is the more complex ability to retrieve *relevant* previous examinations from the archive on a patient scheduled for a new examination. They can be sent to the workstation where the new examination is expected to be routed. While prefetching uses the (auto-) routing path of the scheduled examination, that is the only relationship between these two functions.

Depending on the work environment, autorouting may be based on organ system, examination modality, patient or physician location, radiologist name, etc. Routing paths may even change by time of day or day of week (weekend). The portions of the database schema that define the autorouting must be appropriately site-configurable with no custom programming by the vendor.

Other than hanging protocols (discussed below), it appears that no other folder manager concept has been as difficult to successfully implement as prefetching. Yet any experienced file-room technician (or radiologist) could easily verbalize the types of previous examinations considered relevant for any given new examination. A brute force solution would be to create a large square matrix of examination types or examination mnemonics for a given department (often several hundred in number), with the new or current examination along one axis and the previous ones on the other. The elements relating relevant previous examinations could be flagged and automatically pulled electronically. In practice, this is very difficult to implement.

A better solution might be to realize that these several hundred examination types or mnemonics can be grouped into a much smaller number of categories or metagroups, such as gastrointestinal tract, abdomen, chest, for the purposes of prefetching. These metagroups can be defined in a table of examination mnemonics that maps a particular mnemonic to a metagroup or groups, and vice versa. This table is used to effect the prefetch rules of relevance. A given examination may relate to several prefetch categories, as shown by the example in Fig 3. By constructing the prefetch table in this way, it obeys the fourth normal form of a relational database and facilitates the cross-referencing or searching. The fourth data normalization rule says that no table may contain 2 or more one-to-many (1:M) or many-to-many (M:M) relationships that are not directly related. Ignoring this rule creates a situation known as a multivalued dependency.[6] Preferences for a particular institution should be configurable at each site, including the number of previous examinations to prefetch and how far back in the patient's history imaging examinations should be pulled. In addition, the system should be capable of tagging the pertinent positive findings from prior tests that would be included in the "previous examinations to be pulled."

Figure 3 shows a prefetch metatable example of a CT chest examination as the scheduled new study, with the reason for examination as "suspicion of esophageal carcinoma." This examination type falls under the "chest" prefetch category. Through the prefetch metatable, examination mnemonic, prefetch category, and study table links, the relevant previous examinations to be pulled, if they exist for this patient, include other CT chest examinations, chest x-rays, and esophagrams.

Prefetch SQL Statement

Implementing the prefetch metatable in database SQL involves a many-to-many fetch category strategy. The following general SQL statement carries out the prefetch scheme described above:

```
select a.AccNum, a.ExamMNE from Study_Table a,
  Prefetch_Table b, Prefetch_Table c where a.MRN
  = Pt_IDNumMRN and b.ExamMNE = "exam"
```

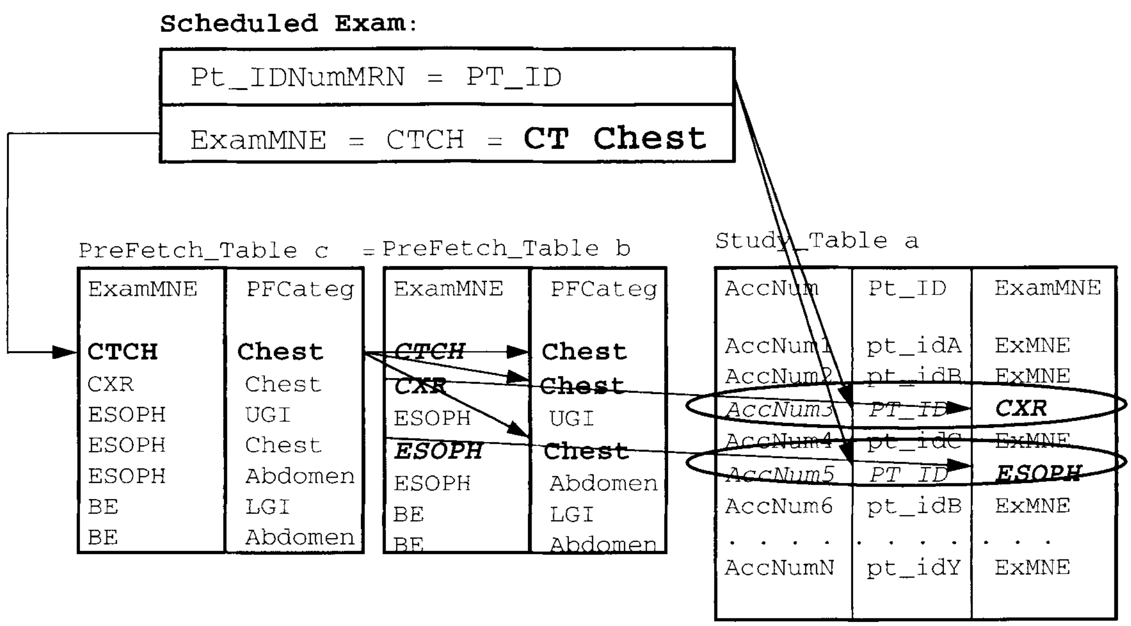


Fig 3. Example of a prefetch metatable with targets circled for a CT chest as the scheduled examination.

and [(c.PFCateg 1 = b. PFCateg 1) or

(c. PFCateg 2 = b. PFCateg 2) or

(c. PFCateg 1 = b. PFCateg 2) or

(c. PFCateg 2 = b. PFCateg 1)] and

a.ExamMNE = c.ExmMNE

where "a" denotes the current examination such that "a.AccNum" is the current examination accession number, "a.ExamMNE" is the current examination mnemonic, "Study_Table a" is the study table for the current study a, "a.MRN" is the current patient's medical record number or patient identification number, "Pt_IDNumMRN," "b" and "c" represent previous examinations, "Prefetch_Table b and c" are identical prefetch tables (for cross-referencing) with examination mnemonics matching "a.ExamMNE" and "PtIDNumMRN" matching "a.MRN," "PFCateg #" denotes the prefetch category set relevant for the current exam "a."

DATABASE ISSUES IN DATA PRESENTATION

Default Views and Worklists

The graphical user interface (GUI) of an image display station must be extremely intuitive and easy to use. Many functions performed by humans in the film-based world, such as hanging the imaging examinations on the light box or alternator in some known examination-specific default view prior to the radiologist's reading them, must now be automated and performed by the PACS software application. Identifying and sorting a day's work of examinations for the image interpreter (ie, providing a worklist) also must be facile and correspond to the way a radiologist works in a given clinical environment. Proper PACS database design directly impacts the usability of an image display station.

The following worklist creation scenario illustrates how a PACS display station database should function. Assume one is working at a display station that can present a default view of today's studies (ie, all unread CT brain examinations from the day), loaded by autorouting to that workstation. To begin reading this set or subset of examinations, one selects (some) studies from the default view, creating a worklist.

A drill-down capability on the worklist studies and patients is required to generate a secondary view of all (relevant) previous studies, autoloaded or otherwise available by fetching from the archive, to the workstation. Drill-down refers to focusing in increasing detail on deeper levels of the database. For example, from a clinical perspective, a patient is the fundamental (coarse) data unit. Once one has focused on a given patient, the next level of detail is the radiology examination history or virtual film jacket. The next level of "increasing granularity," as this is sometimes referred to, is the specific examination, and then the specific relevant previous studies as well as the multiple series of images within a study. To include pertinent previous information, the database view may require a wide time frame (eg, months or years). Yet the selected current studies and patient's list should remain undisturbed, such that other patients and their studies still on the workstation from yesterday or last week should not appear in the list merely because the time frame has enlarged. The manual or automatic selection of new studies should take precedence, and the view of previous ones should be limited to those of patients selected on the worklist, without requiring additional sorting steps by the radiologist user.

Furthermore, once the worklist of current patients and their current and previous studies is created, manually or automatically, it should be persistent and savable for the length of the session (day, week), without having to be recreated. Additionally, integration of the PACS workstation with the dictation system would be useful to prevent duplicate examination interpretation. Loading a study for interpretation can lock the file for that accession number from being loaded by another radiologist for the same purpose. Once an examination has been interpreted, it could be removed from a shared worklist or otherwise locked, to prevent replication of work by another remotely located radiologist involved with the same database.

Worklist Sort SQL Statement

Either a compound embedded select or a self-join SQL statement can accomplish this worklist selection by eliminating the time conflict. The examination directory or patient list has the date of study for the "today" selection, and also contains the individual patients' medical record number (Pt_IDNumMRN). Only the Pt_IDNumMRNs are used for searching, sorting, and listing all relevant previous studies for the selected image worklist subset. The following SQL statement performs this more efficient automatic worklist generation using the embedded select feature:

```
select * from Study_Table where Pt_IDNumMRN
in (select Pt_IDNumMRN from Study_Table where
Exam_Date = "Today")
```

where output values from the inner query are fed as the input values to the outer query. The result of this operation generates a worklist of all patients with most recent examination done "Today," and includes all their relevant previous studies from any date. The same concept could be applied to unread examinations from yesterday or any other specified time. The same result can be achieved using the following self-join SQL statement:

```
select a.* from Study_Table a, Study_Table b
where a.Pt_IDNumMRN = b.Pt_IDNumMRN
and b.Exam_Date = "Today"
```

where "Study_Table b" is scanned for studies done today, and "Study_Table a" joins its rows based on "Study_Table b's" "Pt_IDNumMRN."

Intelligent Hanging Protocols

Providing flexible and easily site-configurable automatic imaging examination soft-copy viewing formats or "hanging protocols" appears to be one of the most difficult issues in workstation design. One solution to this problem lies in the ability to acquire rigorously stable series or sequence naming for all modalities from the imaging device or scanner. With discipline, this probably can be achieved even without bidirectional modality interfaces and DICOM modality worklist capability, by using a menu of editable fixed text. Once the series have predictable names, rules can be recorded within the database to specify format preferences for a given study type, and the presentation and location of particular series or sequences. Hanging protocol schemas should be capable of taking into account the existence of a previous examination, and should be easily site-configurable down to the individual radiologist. An example of a hanging protocol table schema is shown in Fig 4.

Figure 5A diagrams an example hanging protocol for an MR brain study consisting of 4 sequences, on a patient with an existing previous MR brain examination also of 4 sequences. The default viewing mode here is the cine or stack mode for each sequence, (ie, T2 first echo [T2E1], T2 second echo [T2E2], T1 sagittal [T1Sag], and T2 coronal [T2Cor]), with the placement of the sequences as shown. The sequences for the new or current study are placed in the specified order on the left monitor, and the previous study's sequences are placed in the same screen locations on the right monitor.

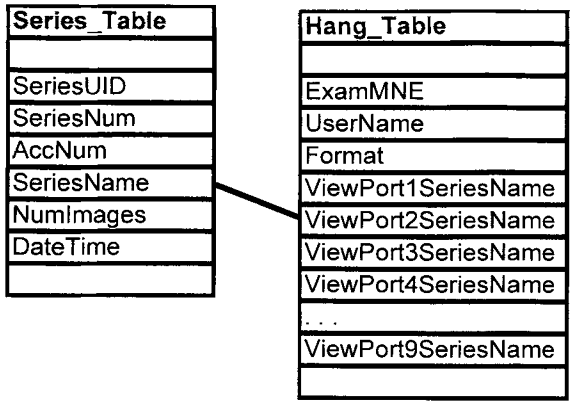


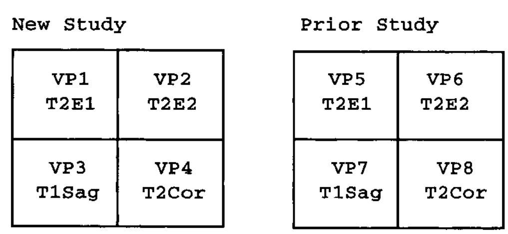
Fig 4. Hanging protocol table schema.

Figure 5B shows an actual display station screen shot of this MR hanging protocol.

Projection Radiography Soft-Copy Display

Any digital projection radiography (ie, computed radiography [CR]) study other than single view, such as a portable chest examination, presents the dilemma of tying the images together in a single series, or storing the images as multiple series but each containing only a single image. The second method is preferred for the following reason. Consider a posterior-anterior (PA) and lateral (Lat) view of the chest. In a single series, the order of display on the workstation monitor is determined by the order in which the plates were placed into the CR reader for processing and subsequent acquisition into the PACS. In most systems, that order is unalterable. Furthermore, it is difficult to name individual images in a series, but as discussed above, easy to specify names for a series (ie, PA, Lat, etc), especially if the series contains only a single image. The ability to name a (single CR image) series facilitates automatic hanging protocols that enable comparison of series with matching names. For example, in evaluating a patient with current and previous 2-view chest examinations, one frequently wishes to compare the current PA

A



Hanging Protocol Table Record

Hanging Protocol Table Record							
MRBrain	WBDMD	Stack/VP8	T2Echo1	T2Echo2	T1Sag	T2Cor	

B

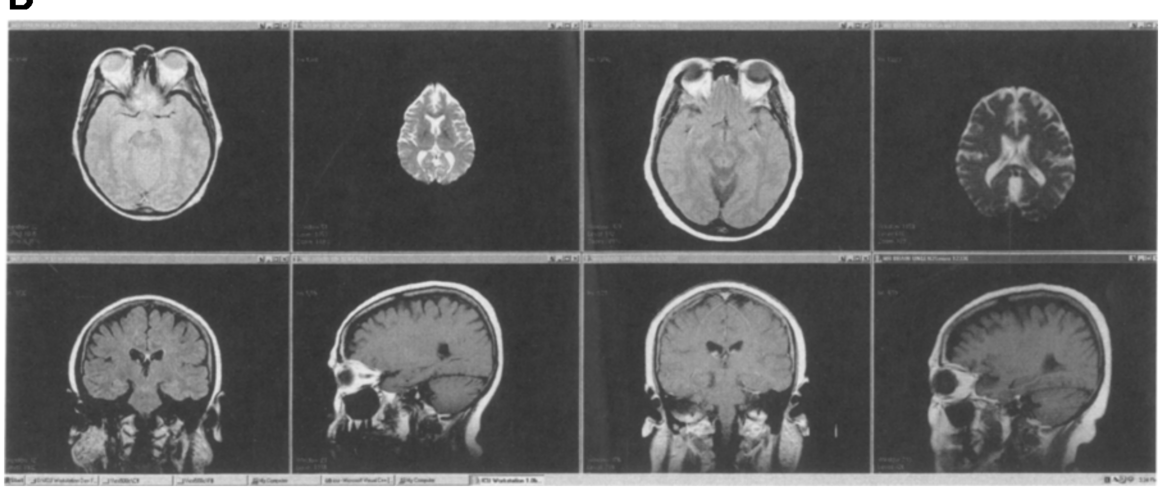


Fig 5. MR new and prior study hanging protocol sample diagram (A) and actual display station screen shot (B).

view to the previous PA view, and the current lateral view to the previous lateral view with proper placement in the display space.

Figure 6A is an example of a hanging protocol for a CR chest study consisting of 2 views (PA and Lat) on a patient with a previous 2-view chest examination. Figure 6B shows an actual display station screen shot of this CR hanging protocol. In this "2-view-with-previous-chest" hanging protocol, the PA views (current and previous) are placed side-by-side on the left monitor, and the lateral views (current and previous) are placed side-by-side on the right monitor. A "blow up" or magnification of any image to the full size should be available instantaneously with a mouse click. The same concept is extensible to orthopedic examinations.

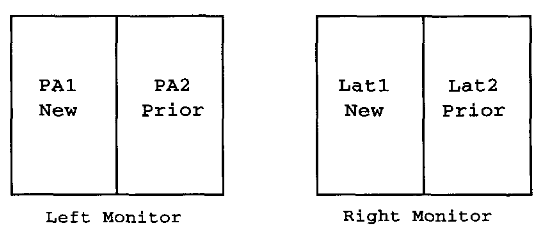
FUTURE ISSUES

Successful implementation of PACS requires intelligent software that emulates file room, reading room, and radiologist procedures and practices, with a minimum of human intervention. In ideal practice, this folder manager or workflow software is rules-based, table-driven, and configurable. Some of the underlying database issues that enable the implementation of these folder manager or intelligent workflow concepts have been outlined.

Further work needs to be done in the area of HIS/RIS/PACS database integration, including more development with the imaging vendors on bidirectional modality worklists and bidirectional quality assurance verification of the HIS/RIS/PACS databases. In time, HIS/RIS/PACS databases may become unified in the virtual sense through intelligent interfacing. Also required are PACS interaction capabilities for end-of-exam notification, and modality control-console design for standardization of series names (ie, MR series descriptors).

Other issues that need to be addressed include series splitting based on procedures requiring an interruption in scanning such as for a breath-hold, and z-axis or slice position sorting for body contrast examinations taken out of anatomic order (ie, a

A



Hanging Protocol Table Record

Hanging Protocol Table Record							
CXRPALat	DEAMD	VP4/NewOld	PA1	PA2	Lat1	Lat2	

B

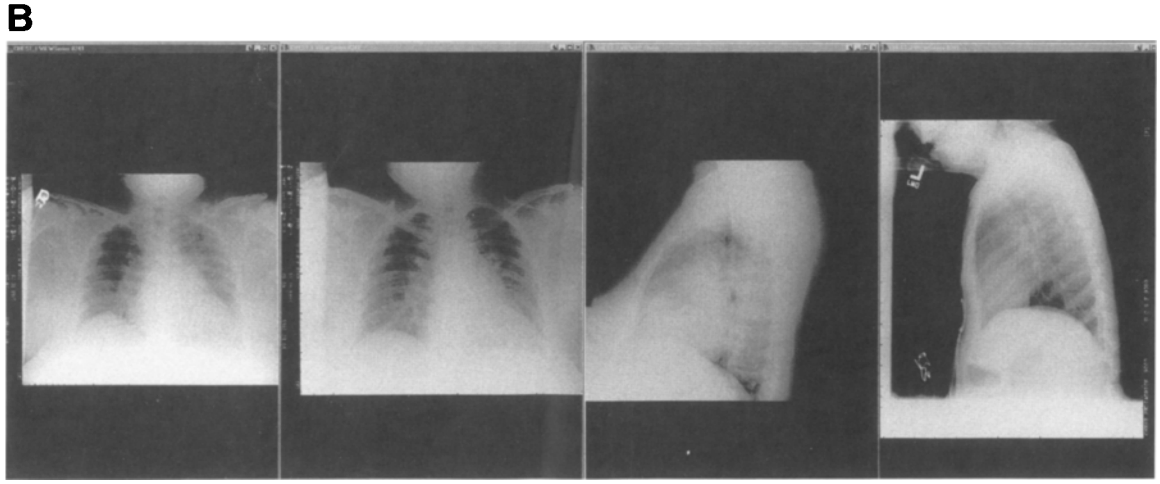


Fig 6. Chest PA/lateral new and previous study hanging protocol sample diagram (A) and actual display station screen shot (B).

chest-abdomen-pelvis examination scanned temporally in the chest-pelvis-abdomen order).

Finally, many issues surrounding the creation of a community master patient index for communication of patient imaging studies and reports across and between health care enterprises need future analysis and development. Database progress in this area may facilitate the realization of the virtual complete electronic medical record.

ACKNOWLEDGMENT

The authors are indebted to Dr Roger A. Bauman for his careful and critical reading of the manuscript, and his thought-provoking suggestions.

REFERENCES

1. Arenson RL, Seshadri SB, Stevens FJ, et al: The overlapping domains and interface between radiology information systems and medical image management systems (PACS), in Lemke HU, Rhodes ML, Jaffee CC, et al (eds): Proceedings of Computer Assisted Radiology. Berlin, Germany, Springer-Verlag, 1987, pp 855-865

2. Seshadri SB, Kishore S, Arenson RL: Software suite for image archiving and retrieval. Radiographics 12:357-363, 1992

3. Feingold E, Seshadri SB, Arenson RL: Folder management on a multimodality PACS display station. SPIE Med Imaging V: PACS Design and Evaluation 1446-3:242-248, 1991

4. Arenson RL, Avrin DE, Wong A, et al: Second generation folder manager for PACS. Proceedings of SCAR '94. Winston-Salem, NC, June 12-15, 1994, pp 601-605

5. HL7 web site at: http://www.hl7.org/

6. Rippy D, Johnson Y: Using Borland Paradox 7. New York, NY, Macmillan, 1999, pp 116-117,