Managing the CT Data Explosion: Initial Experiences of Archiving Volumetric Datasets in a Mini-PACS

Kyoung Ho Lee, M.D.,¹ Hak Jong Lee, M.D.,¹ Jae Hyung Kim, M.D.,¹ Heung Sik Kang, M.D.,¹
Kyung Won Lee, M.D.,¹ Helen Hong, Ph.D.,² Ho Jun Chin, M.D.,³ and Kyoo Seob Ha, M.D.³

Two image datasets (one thick section dataset and another volumetric dataset) were typically reconstructed from each single CT projection data. The volumetric dataset was stored in a mini-PACS with 271-Gb online and 680-Gb nearline storage and routed to radiologists' workstations, whereas the thick section dataset was stored in the main PACS. Over a 5-month sample period, 278 Gb of CT data (8976 examinations) was stored in the main PACS, and 738 Gb of volumetric datasets (6193 examinations) was stored in the mini-PACS. The volumetric datasets formed 32.8% of total data for all modalities (2.20 Tb) in the main PACS and mini-PACS combined. At the end of this period, the volumetric datasets of 1892 and 5162 examinations were kept online and nearline, respectively. Mini-PACS offers an effective method of archiving every volumetric dataset and delivering it to radiologists.

KEY WORDS: Multidetector row computed tomography, volumetric dataset, mini-PACS

Although an awareness of a data explosion crisis has been triggered by the introduction of multidetector-row computed tomography (MDCT),^{1,2} few solutions have been proposed to cope with this overload. The conversion of remaining older scanners to 16-detector-row or higher MDCT and the introduction of the flat panel detector CT will further increase the number of images.³ With the increased availability of volumetric scanners and three-dimensional (3-D) workstations, the next challenge might be to effectively archive and move thin-section datasets for 3-D visualization (we refer to these as volumetric datasets) around an enterprise. Although a limited number of large academic institutions are storing all of these large datasets in their Picture Archiving and Communication System (PACS), this ideal dataflow requires a large-capacity PACS.

Furthermore, unless another thick section dataset is distributed, many reviewing workstations throughout the enterprise will require 3-D visualizing functionalities such as the slab multiplanar reformation (MPR) for reviewing these thin-section datasets, if image noise must be kept at a low level.⁴

Therefore many radiologists today are adopting expedients such as acquiring MDCT images using thin collimation and then reconstructing images for PACS storing using much thicker (ca. 5 mm) sections.⁵ The reconstruction of thinner sections for volumetric datasets can then be performed only for several special studies, such as CT angiography, or on a radiologist's request. Moreover, because these volumetric datasets have a substantial

¹From the Department of Radiology, National University Bundang Hospital, National University College of Medicine, Institute of Radiation Medicine, National University Medical Research Center, 300 Gumi-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-707, Korea.

²From the School of Electrical Engineering and Computer Science, BK21: Information Technology, Seoul National University, San 56-1 Shinlim-dong, Kwanak-gu, Seoul 151-747, Korea.

³From the Department of Medical Informatics, National University Bundang Hospital, 300 Gumi-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-707, Korea.

Correspondence to: Jae Hyung Kim, M.D., Department of Radiology, National University Bundang Hospital, 300 Gumi-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-707, Korea; tel: 82-31-7877609; fax: 82-31-7874011; e-mail: jaehkim@radiol.snu.ac.kr

Copyright © 2005 by SCAR (Society for Computer Applications in Radiology)

Online publication 2 June 2005

doi: 10.1007/s10278-005-5163-z

impact on storage needs and network traffic, they are occasionally sent directly to a dedicated 3-D workstation instead of being stored in PACS.

From our experience of four-detector-row MDCT scanners, volumetric datasets are accessed by only a limited number of medical personnel, and they are scarcely retrieved if once interpreted. With these characteristics in mind, we implemented a mini-PACS with a smaller online storage and fewer concurrent user sessions to exclusively handle volumetric datasets from 16-detector-row CT scanners. In our institution, the volumetric datasets of every CT examination are stored in the mini-PACS and routed (or prepushed) to the relevant radiologist's workstations, and another dataset of thicker section is distributed in the main PACS network for routine clinical use. The purpose of this study was to describe our new dataflow system and to report our initial experiences with this system.

METHODS

This study was conducted at a 900-bed tertiary care academic medical center which opened in May 2003. Workloads in the Department of Radiology and of the PACS were rapidly increasing during the study period. The hospital was equipped with two 16-detector-row CT scanners (Mx8000; Philips Medical Systems, Cleveland, OH) only. PACS (Impax Enterprise; Agfa, Mortsel, Belgium) storage was provided by a redundant array of inexpensive disks (RAID) with an available storage of 4.7 Tb and a robotic Digital Linear Tape (DLT) library with an available storage of 13.5 Tb after mirroring data. In this study, online, nearline, and offline storage indicate RAID, the robotic DLT library, and older DLTs removed from

the robotic system, respectively. Image data were typically compressed in the range 2.5:1 to 3:1 in a lossless manner, stored online, and then archived in the mirrored nearline storage at night. The system can delete older studies in online storage using an algorithm based on creation date, last access time, and access number to maintain space for new studies. However, this deletion functionality has not been used, and we have not activated prefetch yet because we have been gradually increasing the online storage to keep all studies online. The dataflow in our new system is summarized in Figure 1. Two image datasets (one thick section dataset and another volumetric dataset) were typically reconstructed from a single CT projection data. The thick section dataset was generated mostly with a thickness of 4–5 mm with a 20–25% overlap and distributed in the main PACS network. The volumetric dataset was generated with a thickness of 0.8 or 2 mm with a 50% overlap, which corresponds to a minimum detector collimation of 0.75 or 1.5 mm for our CT scanners. This dataset was transferred through a 100-Mbps fast Ethernet network to a dedicated mini-PACS (Impax Basix; Agfa). The original development purpose of the mini-PACS is being more affordable for smaller hospitals. The mini-PACS ran on a server (Compaq ProLiant ML570; Houston, TX) with four processors, attached with a 271-Gb RAID and a robotic DLT library (L40; Storagetek, Louisville, CO) with an available storage of 680 Gb after data mirroring. The compression, archiving, and persistence models used for image data were the same as those used in the main PACS. No prefetch rule was set in the mini-PACS. Older tape was regularly removed from the robotic DLT system and kept offline for permanent storage. CT data acquisition, image reconstruction, and transmission were performed automatically with a simple operation by a technologist. All of these processes, including transferring the thicker section dataset and volumetric dataset to the main and the mini-PACSs, respectively, were programmed into the CT scanner protocol. Scanners automatically reconstructed and transferred images in a stream while further scanning proceeded.

The volumetric dataset was routed, or prepushed, in an uncompressed state to one or more 3-D workstation(s) (Rapidia, Infinitt, Seoul, South Korea, and MxView, Philips

Fig 1. Proposed CT dataflow. The volumetric dataset is routinely reconstructed and stored in a dedicated mini-PACS, which has a smaller online storage and fewer concurrent user sessions. The volumetric dataset is routed to the relevant 3-D workstation(s) (dotted arrows), whereas another axial dataset of thicker section is distributed in the main PACS network.

Medical Systems) by following predefined rules based on study classifications by radiologist specialty. For example, an abdomen and pelvis CT was always routed to five workstations in the CT unit ($n = 1$), the reading rooms for gastrointestinal ($n = 2$) and urogenital radiologists ($n = 1$), and the gastrointestinal radiologist's office ($n = 1$). Digital Imaging and Communications in Medicine (DICOM) query and retrieval was supported for exceptional cases. Twenty-eight workstations featuring 3-D visualization such as MPR, maximum intensity projection, volume rendering and fly-through, and three standard diagnostic display stations (CS5000; Agfa) were connected to the mini-PACS through a gigabit or 100-Mbps fast Ethernet network. In the reading rooms, eight of these workstations with dual processors connected to the main PACS simultaneously ran 3-D visualization software (Rapidia) and the PACS viewer program (DS3000; Agfa). Other dedicated 3-D workstations (including four MxView workstations) were located at the CT unit ($n = 3$), the angiography unit ($n = 2$), reading rooms ($n = 4$), conference room ($n = 1$), research laboratories ($n = 2$), radiologists' offices ($n = 5$), neurovascular surgeons' offices ($n = 2$), and one orthopedist's office ($n = 1$). Twenty-four workstations (Rapidia) automati-

cally deleted older studies from local disks by using user-defined thresholds in either of maximum persistence period for a study or minimum percentage for free disk space to maintain space for new studies. Each of the other four workstations (MxView) had 126 Gb of local disc space for volumetric datasets, and older studies in workstations had to be manually deleted to allow new studies to be received.

Technologists routinely performed standardized 3-D postprocessing using the volumetric datasets at the three dedicated 3-D workstations (two MxView and one Rapidia workstation) or main scanner consoles. For example, a set of coronal MPR images was generated at a thickness of 5 mm and a 20% overlap to examine chest, abdomen, or pelvis. These 3-D images were routinely stored as a new series attached to the examination in the main PACS. In the same manner, the radiologists added 3-D images during their interpretation. Our imaging and archiving protocol for each body part is summarized in Table 1.

During the first months after installing our first 16-detector-row CT scanner in April 2003, we gradually took up the clinical work with this new dataflow, whereas the major part of our examinations still used conventional dataflow—recon-

Table 1. Standardized 16-detector-row CT imaging and archiving protocols

Area	Main PACS			Mini-PACS
	Axial datasets ^a	Standardized 3-D image ^b		Volumetric dataset ^a
		MPR	Others	
Body				
CT colography	5/4	Coronal, 5/4	VR, fly-through	2/1
Rectum	5/4	Coronal and sagittal, 5/4	—	2/1
Others	5/4	Coronal, 5/4	—	2/1
Musculoskeletal				
Spine				
Intervertebral disk	3/3 (sequential)	—	—	—
3-D	4/3	Coronal and sagittal, 4/3	VR	2/1
Joint	4/3	Coronal and sagittal, 4/3	VR	2/1
Cardiovascular				
CT angiography	5/4	Coronal, 5/4	VR, MIP	2/1
Heart	5/3	Short axis, four-chamber, and two-chamber views, 5/3	VR, quantitative vessel analysis	0.8/0.4 ^c
Brain, head, and neck				
Brain	6/6 (sequential)	—	—	—
Head and neck	4/3	Coronal, 4/3	—	2/1
	0.75/0.75 (sequential, axial, and coronal)	—	—	—
Temporal bone	3/3 (sequential, axial, and coronal)	—	—	—
Paranasal sinus	4/3	Coronal, 4/3	VR	2/1
Facial bone	4/3	Coronal, 4/3	VR	2/1

Standardized 3-D postprocessing is routinely performed using volumetric datasets. Numbers separated by a virgule indicate slice thickness and reconstruction interval, respectively. "—" indicates that the given image dataset was not reconstructed and, therefore, not archived.

MIP = Maximum intensity projection; VR = volume rendering.

^aImages were reconstructed from the same CT projection data.

^bPostprocessing using the volumetric dataset.

^cFive to ten volumetric datasets were reconstructed at different cardiac phases.

structuring only axial datasets of thick-section images and distributing these images in the main PACS network. In September 2003, we switched to the new dataflow system and abandoned the old dataflow.

To estimate the performance of the new dataflow system, we analyzed the transfer time from the CT scanners to one of the diagnostic display stations via mini-PACS routing. This analysis was performed using the volumetric datasets of 32 abdomen studies, which were performed over 24 hr on a weekday at the end of February 2004. Data transfer time was defined as the time interval between the acquisition of the final image and the arrival of this image at the diagnostic display station. The acquisition time of each reconstructed image was recorded by the DICOM header, and the arrival time of each image was recorded using an auditing software (Infinit) in the diagnostic display station, which was connected to the gigabit network. The system clocks of the CT scanners and of the diagnostic display station were synchronized for these measurements. For the same sample examinations, we also measured the times required for DICOM retrieval from online and nearline mini-PACS storage. These measurements were performed at the same diagnostic display station at 0900–1300 on another weekday at the end of February 2004.

Using the new dataflow system, a CT examination stored in the main PACS included thick-section axial dataset, standardized 3-D images produced by technologists, 3-D images added by a radiologist, and images for scan planning (scout and bolus-tracking images), whereas the mini-PACS contained only thin-section volumetric dataset. To measure the respective storage needs of each dataset type and to estimate volumetric dataset utilization, we measured the number of images and the data volume for each dataset type in each examination. These results are reported as medians, 25% quartiles, and 75% quartiles because the distributions of data were highly asymmetric. The impact on the storage need of each dataset

type was assessed using percentage data volume, which was defined as the data volume of the given dataset divided by the total CT data volume in the main PACS and mini-PACS. The frequency of access to routed volumetric datasets was also measured by reviewing the log files of the auditing software installed in 11 3-D workstations (Rapidia) in reading rooms. For practical reasons, these analyses were performed for examinations conducted during the first 2 weeks of February 2004.

To evaluate the total storage need of our new dataflow system, we analyzed the cumulative number of CT examinations, the cumulative CT data volume, and the space used in online and nearline storage for CT examinations. This analysis was performed for the main PACS and the mini-PACS by measuring these data at the end of every month between September 2003 and February 2004. We also analyzed the cumulative total data volume for all imaging modalities in the main PACS during the same period.

RESULTS

For the 32 sample abdomen studies, which were used to measure data transfer times, all volumetric datasets were routed without fail from the CT scanners to the diagnostic display station. For these sample examinations, the number of images ranged from 462 to 1647 (mean $\pm$ SD, 710 ± 277 ; median, 562), and the data volume ranged from 236 to 847 (357 ± 144 ; 274) Mb in an uncompressed state. The measured data transfer time for an examination ranged 7.5 to 22.6 min. For the same sample examinations, the time required to

Fig 2. Box and whisker plot of the data transfer time and the retrieval time of the volumetric datasets from 32 sample examinations. "Routing" indicates the data transfer time from the CT scanners to a diagnostic display station through the mini-PACS. "Retrieve online" and "Retrieve nearline" indicate the times to retrieve from online and nearline storage in the mini-PACS. The middle lines of boxes show medians. The upper and lower margins of the boxes show upper and lower quartiles, and the upper and lower ends of the vertical lines show upper and lower extremes, respectively.

Table 2. Statistics of image numbers and stored data volumes for each dataset type created during the 2-week sample period

	All examinations	Examinations with volumetric datasets	Examinations without volumetric datasets
Number of examinations	867	604	263
Main PACS			
Total data volume (Gb)	26.8	24.2	2.6
Preparation scan ^{a,b}			
Number of images	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)	2.0 (1.0, 3.0)
Data volume (Mb)	0.4 (0.3, 0.7)	0.5 (0.3, 0.8)	0.3 (0.2, 0.5)
Axial dataset ^a			
Number of images	96.0 (44.3, 130)	120.0 (92.0, 169.3)	27.0 (24.0, 78.0)
Data volume (Mb)	18.3 (9.8, 23.4)	21.2 (17.5, 27.8)	4.0 (3.7, 15.0)
Standardized 3-D ^{a,c}			
Number of images	45.0 (0, 50.0)	45.0 (42.0, 60.0)	NA
Data volume (Mb)	8.7 (0, 10.7)	9.5 (7.9, 11.1)	NA
Additional 3-D ^{a,d}			
Number of images	0 (0, 0)	0 (0, 0)	NA
Data volume (Mb)	0 (0, 0)	0 (0, 0)	NA
Total ^a			
Number of images	150.0 (70, 188.0)	169.0 (142.0, 257.5)	29.0 (25.0, 79.0)
Data volume (Mb)	29.0 (14.4, 35.0)	31.5 (28.1, 46.6)	4.4 (4.0, 15.3)
Mini-PACS			
Total data volume (Gb)	NA	78.1	NA
Volumetric dataset ^a			
Number of images	NA	539.0 (362.5, 738.3)	NA
Data volume (Mb)	NA	99.4 (76.2, 137.0)	NA

NA = Not applicable.

^aData shown are medians, and the data in parentheses are 25 and 75% quartiles.^bScout and bolus-tracking images.^cRoutinely performed by technologists.^dAdded by radiologists.

retrieve a volumetric dataset ranged from 2.8 to 7.6 min from online storage and from 3.5 to 11.5 min from nearline storage (Fig 2).

During the first 2 weeks of February 2004, 867 CT examinations totaling 26.8 Gb (in a compressed state) were stored in the main PACS. These examinations were from the following areas: 300 brain, head, and neck, 251 chest, 209 abdomen and pelvis, 53 musculoskeletal, 40 cardiovascular, and 14 combined chest, abdomen, and pelvis. Of these 867 examinations, the volumetric datasets of 604 examinations (78.1 Gb in a compressed state) were stored in the mini-PACS. For the remaining 263 examinations [171 brain, 37 temporal bone, 30 paranasal sinus, 22 spine (intervertebral disk), and 3 pulmonary nodule biopsy], thin-section volumetric datasets were not reconstructed from the CT projection data. Table 2 summarizes the numbers of images and the data volumes for each dataset type. Table 3 summarizes the impact on storage need of each dataset type for the 867 examinations, presented in percentage data

Table 3. The impact on the storage needs of dataset types for 867 CT examinations performed during the 2-week sample period

	Data volume (Mb)	Impact on storage need ^a (%)
Main PACS		
Preparation scan ^b	0.5	0.5
Axial dataset	16.6	15.9
Standardized 3-D ^c	7.3	7.0
Additional 3-D ^d	2.4	2.3
Total	26.8	25.6
Mini-PACS		
Volumetric dataset	78.1	74.4
Total	104.9	100.0

^aDefined as the data volume of the given dataset divided by the total CT data volume in the main PACS and mini-PACS.^bScout and bolus-tracking images.^cRoutinely performed by technologists.^dAdded by radiologists.

Fig 3. Monthly changes in the cumulative data volume in the main PACS and mini-PACS, and the spaces used in online and nearline storage of the mini-PACS, during the 5-month period from October 2003 to February 2004.

volume. In 148 (24.5%) of the 604 examinations stored in the mini-PACS, the routed volumetric dataset was accessed at least once by radiologists. In 110 (18.2%) examinations, the image data contained 1–156 (median, 20) 3-D images added by a radiologist.

Figure 3 summarizes monthly changes in the cumulative data volume (in a compressed state) in the main PACS and mini-PACS, and the used space in the online and nearline storage of the mini-PACS, during the 5 months from October 2003 to February 2004. During this 5-month period, 1.48 Tb of image data from various modalities was stored in the main PACS. Of these image data, 278 Gb was from 8976 CT examinations. During the same period, volumetric datasets of 738 Gb from 6193 examinations were stored in the mini-PACS. These volumetric datasets formed 72.6% of the total CT data (0.99 Tb) and 32.8% of total data for all modalities (2.20 Tb) in the main PACS and mini-PACS combined. The online and nearline storages of the mini-PACS reached a steady state in October 2003 and January 2004, respectively. At the end of the 5-month period, all image data in the main PACS were kept online, whereas in the mini-PACS, 1892 examinations of

245 Gb were in online storage and 5162 examinations of 650 Gb were in nearline storage. Because we performed approximately 1795 CT studies per month during the 5-month period, the persistence periods for a volumetric dataset in online and nearline storage were estimated to be 1.1 and 2.9 months, respectively.

DISCUSSION

To the best of our knowledge, there is no legal requirement that radiologists should archive, distribute, and interpret every volumetric dataset; policies vary between radiologists and institutions. In our experience, once a volumetric dataset has been initially interpreted or postprocessed, it is rarely retrieved because many referring physicians and even radiologists prefer reviewing thick-section images and saved 3-D images during patient follow up. Therefore it might be attractive to bring volumetric dataset nearline and then off-line earlier than the other components in an examination. However, to the best of our knowledge, such intelligence is not always provided by a

single commercial PACS, including ours. In this study, we propose an uncomplicated solution, that is, by implementing the mini-PACS with a smaller online storage.

In our experience, volumetric datasets are usually accessed by a limited number of medical personnel using a high bandwidth network and with a 3-D workstation, and many referring physicians at busy outpatient clinics view these source data as troublesome. Therefore a mini-PACS with fewer concurrent user sessions may be sufficient to distribute volumetric datasets.

For our new dataflow system, the mini-PACS with smaller online storage and fewer concurrent user sessions took over the volumetric datasets, thus freeing the main PACS from the burden of these data, which accounted for 72.6% of all CT data and 32.8% of the total data for all modalities. By effecting this change, we believe that unnecessary expansion of entire PACS can be avoided with respect to covering increasing volumetric dataset load. Once online storage of the mini-PACS had reached a steady state, DLT could be used, instead of more expensive online storage of the main PACS,⁶ as an additional storage media for volumetric datasets. By extrapolating increase storage trends in Figure 3, it is inferable that this cost-saving will eventually compensate for our initial mini-PACS investment. However, we could not verify this by performing a formal cost analysis because our new dataflow has been used for a relatively short period of time.

In our new dataflow, volumetric datasets were prepushed to 3-D workstations. However, because it was difficult to anticipate perfectly where specifically a given volumetric dataset would be reviewed, many studies were ruled to be sent to multiple workstations, and many of these rules were changed by radiologists' requests during the study period. Although the routing process required longer data transfer times than the DICOM retrieval, this routing proceeded in the background while radiologists were interpreting other studies. Therefore most studies awaiting initial interpretation, if not all, were already in the local disk, thus eliminating data transfer delay.

There are undoubtedly many alternative approaches to effectively managing volumetric datasets from MDCT scanners, and we make no claim that our method is the best. The limitations of our new dataflow system are as follows. First, it re-

quires additional hardware for the mini-PACS, although with smaller capacity. The cost effectiveness of the proposed system should be analyzed by comparing it with a single enterprise PACS where volumetric data are handled effectively using separate rules. Second, access to the volumetric datasets of prior studies for comparison is limited because of the smaller storage. Third, security issues are raised by storing multiple copies of volumetric datasets around a network. Fourth, we have not integrated 3-D postprocessing tasks into our workflow by introducing Post-Processing Workflow Integration Profile in Integrating the Healthcare Enterprise Radiology Integration Profiles⁷ or Clinical Context Object Workgroup interfaces.⁸

CONCLUSION

Despite its limitations, however, our new dataflow offers an effective method of archiving every volumetric dataset and delivering it to radiologists, at least during the transitional period until a more ideal system becomes commercially available and affordable.

ACKNOWLEDGMENTS

This study was supported by a grant from the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (03-PJ1-PG3-21900-0002). We thank Chang Min Dae, R.T. in our CT unit, Sang Tae Kim, R.T. in Department of Medical Informatics, Chang Goo Baek and JinSeo Kim in Agfa Korea, and Dong Gyun Cheong, Ph.D. in Infinitt Co., Ltd., for their technical assistance with the system integration and data collection.

REFERENCES

1. Rubin GD: Data explosion: the challenge of multi-detector-row CT. *Eur J Radiol* 36:74-80, 2000
2. Tamm EP, Thompson S, Venable SL, McEnery K: Impact of multislice CT on PACS resources. *J Digit Imaging* 15(Suppl 1):96-101, 2002
3. Kalender WA: The use of flat-panel detectors for CT imaging. *Radiologie* 43:379-387, 2003
4. Prokop M: Image processing and display techniques. In: Prokop M, Galanski M (eds). *Spiral and Multislice Computed Tomography of the Body*. Stuttgart, Germany: Georg Thieme Verlag, 2003, pp 49-50
5. Siegel EL: Interpretation strategies for large imaging datasets. In: Reiner BI, Siegel EL (eds). *SCAR University*

Syllabus at the 2004 Annual Meeting. Great Falls, VA: Society for Computer Applications in Radiology, 2002, pp 104–105

6. Chunn T, Honeyman J: Storage and database. In: Kim Y, Horii SC (eds). *Handbook of Medical Imaging Volume 3 Display and PACS*. Bellingham, WA: The International Society for Optical Engineering, 2000, pp 369–378

7. Channin DS: Integrating the healthcare enterprise: a primer. Part 6: the fellowship of IHE: year 4 additions and extensions. *Radiographics* 22:1555–1560, 2002

8. Health Level Seven: Clinical Context Management Specification (CCOW) Version 1.4. Ann Arbor, MI: Health Level Seven, 2002, pp 369–378