# Journal of Digital Imaging

## Collecting 48,000 CT Exams for the Lung Screening Study of the National Lung Screening Trial

Kenneth W. Clark,<sup>1</sup> David S. Gierada,<sup>1</sup> Guillermo Marquez,<sup>2</sup> Stephen M. Moore,<sup>1</sup> David R. Maffitt,<sup>1</sup> Joan D. Moulton,<sup>1</sup> Mary A. Wolfsberger,<sup>1</sup> Paul Koppel,<sup>1</sup> Stanley R. Phillips,<sup>1</sup> and Fred W. Prior<sup>1</sup>

From 2002-2004, the Lung Screening Study (LSS) of the National Lung Screening Trial (NLST) enrolled 34,614 participants, aged 55-74 years, at increased risk for lung cancer due to heavy cigarette smoking. Participants, randomized to standard chest X-ray (CXR) or computed tomography (CT) arms at ten screening centers, received up to three imaging screens for lung cancer at annual intervals. Participant medical histories and radiologist-interpreted screening results were transmitted to the LSS coordinating center, while all images were retained at local screening centers. From 2005-2007, all CT exams were uniformly de-identified and delivered to a central repository, the CT Image Library (CTIL), on external hard drives (94%) or CD/DVD (5.9%), or over a secure Internet connection (0.1%). Of 48,723 CT screens performed, only 176 (0.3%) were unavailable (lost, corrupted, compressed) while 48,547 (99.7%) were delivered to the CTIL. Described here is the experience organizing, implementing, and adapting the clinical-trial workflow surrounding the image retrieval, de-identification, delivery, and archiving of available LSS-NLST CT exams for the CTIL, together with the quality assurance procedures associated with those collection tasks. This collection of CT exams, obtained in a specific, well-defined participant population under a common protocol at evenly spaced intervals, and its attending demographic and clinical information, are now available to lung-disease investigators and developers of computer-aided-diagnosis algorithms. The approach to large scale, multi-center trial CT image collection detailed here may serve as a useful model, while the experience reported should be valuable in the planning and execution of future equivalent endeavors.

KEY WORDS: Cancer detection, chest CT, clinical trial, computed tomography, de-identification, lung diseases, digital image management, image database, image libraries, national lung screening trial, lung screening study, CT image library

#### BACKGROUND

T he National Lung Screening Trial (NLST) aims to compare the effectiveness of two screening tests, low-dose spiral CT scan and CXR, with respect to their impact on lung-cancerspecific mortality, in persons who are at high risk for developing lung cancer due to their age (55– 74 years) and heavy smoking history (at least 30 pack years). The trial is sponsored by the National Cancer Institute (NCI) and conducted under a harmonized protocol within two separate administrative organizations: the Lung Screening Study (LSS)<sup>1</sup> and the American College of Radiology Imaging Network (ACRIN)<sup>2</sup>. The trial is currently in the post-screening, clinical follow-up phase, with final outcomes analysis planned for 2009.

Between September 4, 2002 and April 26, 2004, the LSS enrolled 34,614 participants who were randomized to CXR or CT imaging arms. Three serial screening exams {T0 (Time 0, or baseline),

Copyright @ 2008 by Society for Imaging Informatics in Medicine

*Online publication 6 September 2008 doi: 10.1007/s10278-008-9145-9* 

<sup>&</sup>lt;sup>1</sup>From the Mallinckrodt Institute of Radiology, Washington University, 510 South Kingshighway, Saint Louis, MO 63110, USA.

<sup>&</sup>lt;sup>2</sup>From the National Cancer Institute, Early Detection Research Group, Bethesda, MD USA.

Correspondence to: Kenneth W. Clark, Mallinckrodt Institute of Radiology, Washington University, 510 South Kingshighway, Saint Louis, MO 63110, USA; tel: +314-362-6965; fax: +314-362-6971; e-mail: clarkk@mir.wustl.edu

T1, T2} were performed at approximately annual intervals. The final screening exam was performed January 16, 2007. Participants were enrolled through and screened at ten LSS Screening Centers (SCs) and two operationally-independent satellite centers, considered here as SCs (Appendix A). The SCs operated within the screening centers of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial Network<sup>3</sup>. In order to rapidly achieve enrollment targets, most SCs coordinated recruitment and screening through multiple local and regional medical centers.

Screening exams were performed, interpreted, and archived at local SCs. All CT exams and 62.6% of CXRs (35.3% CR; 27.3% DR) were archived digitally, while 37.4% CXRs were screen films. Screening results (radiologist interpretations), participant demographic and baseline health data, and medical follow-up information were forwarded to the LSS coordinating center, Westat (Rockville, MD, USA), an independent research firm, contracted by NCI to provide coordinating and statistical services for the LSS network.

Throughout the screening period, the Electronic Radiology Laboratory (ERL) of the Mallinckrodt Institute of Radiology (MIR), Washington University School of Medicine (Saint Louis, Missouri, USA) managed the imaging Quality-Assurance Coordinating Center for the LSS network<sup>4</sup>. Based on this experience, NCI contracted with ERL/MIR, in 2004, to assemble and administer an LSS-NLST CT Image Library (CTIL), consisting of exact digital copies of all previously interpreted LSS CT exams. Institutional Review Board approval to conduct the NLST, with informed consent from all participants to store de-identified CT images in a central database for use in future research, was obtained at all screening centers. Data-use agreements were established with each SC that allowed transfer of deidentified images to the CTIL. With approximately half (17,309) of the participants randomized to the CT arm, the CTIL would need to accommodate a maximum of 51,927 CT exams. A total 48,723 CT screens were actually performed. The number performed varied among LSS SCs (average 4,061; stdev 2,229; range 848–8,565) as did (a) the number of radiologists interpreting those screens (9.9; 6.5; 2-21) and (b) the number of exams per radiologist (686; 675; 103-2,141). Delivery of CT screening exams to the CTIL began in January 2005, more than 2 years after screening began.

The CTIL archive was created as a resource for use by NLST investigators, lung-disease clinical researchers, and software developers of computer-aided diagnosis algorithms. The CT exams were obtained in a specific, well-defined participant population whose members were scanned under a common protocol at evenly-spaced intervals. The exams are associated with a separate database, managed by Westat, of baseline and follow-up clinical information, features expected to enhance their value to investigators.

The hardware and software infrastructure of the CTIL was detailed shortly after the collection of CT examinations began<sup>5</sup>. Here, we describe our experience organizing, implementing, and adapting the clinical-trial workflow surrounding the image retrieval, de-identification, delivery, and archiving of copies of all available LSS-NLST CT exams for the CTIL, together with the quality-assurance procedures associated with those collection tasks and the problems encountered along the way. This may serve to aid in planning similar image collection efforts in other multicenter clinical trials.

#### METHODS

#### Screening-center Activities

*CT Scanners*. Prior to the commencement of screening activities, NLST medical physicists, from both ACRIN and LSS, determined target scanner settings and allowable latitudes across scanners in order to formulate a consistent CT screening protocol<sup>6</sup>. These same physicists then directed the collection of periodic scanner calibration data to ensure consistent screening throughout the trial.

LSS SCs used multidetector CT scanners (various vendors) available to them at their medical centers to perform participant screens (Table 1). Because of the vendor–scanner variation, some scanner parameters relevant to NLST-protocol adherence could be obtained directly from DICOM image headers while others were obtained indirectly based on other information in the image headers. Those parameters saved in the CTIL management database are reported in Appendix B, along with details of direct or indirect capture.

*Enrollment and Screening.* At enrollment, each participant was assigned an NLST Participant Identifier (PID) that included a two-digit SC-number

Table 1. CT Scanners by Vendor and LSS Screening Center

SC	Vendor				
	GE	Philips	Siemens	Toshiba	
A	x				
В	x	x	x		
С			х		
D			х		
E	x				
F	x		х		
G		х		х	
Н	x				
I		х	х		
J	x			х	
К		х	х		
L	х				

prefix. The PID, participant gender, date of birth (DOB), and NLST-relevant medical history were sent to Westat as part of the enrollment process. At each screening, the PID, exam date, and visit number were likewise sent to Westat. The visit number distinguished multiple screens, in the same screening year, when a subsequent imaging exam was obtained to replace a prior one of poor diagnostic quality, as determined by the interpreting radiologist. The incidence of multiple visits was quite small: 151 of 48,723 screens (0.3%). SCs were not required, though three chose, to insert NLST PIDs or other de-identifying characters in place of local medicalcenter patient IDs in the DICOM image headers at the time of screening. Because exams could not be guaranteed to contain NLST PIDS, all exams were uniformly de-identified, with NLST PIDs inserted, before the exams were sent to the CTIL.

#### Exam Collection Preparation

*Hardware and Software.* ERL/MIR provided each SC with a Dell Inc. (Round Rock, TX, USA) Inspiron 1150 laptop computer plus DVD writer and a Maxtor Corporation (Milpitas, CA, USA) 250-GByte external hard drive (XHD). To facilitate the collection process, the laptop was pre-loaded with DICOM communications software, virtual private network software, de-identification software, a graphical user interface (GUI) application called the Clinical Studies Workstation (CSW), and a user's guide. The de-identification software and GUI<sup>7</sup> were customized for CTIL collection<sup>5</sup>.

PACS to Laptop Communications. The DICOM communications software loaded to the laptops

was written at the ERL and has been used in other clinical trials and a variety of ERL-based research projects. The underlying communication software uses the MIR Central Test Node software<sup>7</sup> that has been widely tested in the industry over the last 15 years. The SCs were required to work with network and Picture Archive and Communications System (PACS) teams at their institutions to determine (a) physical placement of the laptop, (b) the appropriate network common access of the laptop and the PACS, (c) resolution of any firewall issues, (d) whether the laptop would "query" the PACS for image studies or the PACS would "push" studies to the laptop or both (the choice was dictated primarily by PACS policy at the SC but the laptop DICOM communications software was configured for all choices). In any case, the PACS was configured-to show the laptop as a legitimate destination. The user's guide supplied with the laptop included instructions for activating and using the DICOM communications software. For the most part, SCs were able to test their laptop-PACS configurations without assistance from ERL; minor problems were resolved by telephone. In all instances but one, SCs were able to transmit images from their local PACS to the CSW using the standard DICOM protocol; at one SC, the exams were loaded from compact disks.

*Workforce.* SCs budgeted the number of full-time equivalent (FTE) personnel for CTIL work based on numbers of CT exams to be collected and delivered to the CTIL as well as the period in which the work would be performed. The personnel responsible for retrieving CT exams from local archives, performing the de-identification, and copying and delivering the images to the CTIL were information-technology specialists (at seven of 12 SCs), reporting to SC coordinators or the coordinators themselves (at five of 12 SCs).

*Training.* SC experience in de-identifying images as part of the ongoing imaging QA work, plus a detailed user's guide, obviated the need for the onsite training for CTIL collection offered by CTIL personnel. Each SC became familiar with the CTIL submission procedures by first delivering a small number (one to five) of exams to the CTIL. Problems that surfaced were resolved with telephone dialog and email exchanges, leading, in turn, to minor revisions to the user's guide. Exam Retrieval, Matching, and De-identification. While initial screens were performed late 2002, CTIL collection did not begin until early 2005. At that point, Westat provided each SC with a list of screens thus far performed; this list included, for each exam: PID, DOB, gender, exam date, and visit number. The SC loaded this list to the laptop, and the list was used as a "Matching-List" in the de-identification process. After this initial list, Westat sent to each SC, on a monthly basis, a list of new screens performed since the last list was provided. The new list was added to the prior lists on the laptop to maintain a cumulative Matching-List. As previously noted, exams reaching the laptop contained patient names or NLST PIDs, as well as other local identifiers, according to SC practice. Regardless, all exams were uniformly deidentified at the laptops. Two SCs maintained their own LSS-specific image archive (one of these simply used compact discs) instead of their medical centers' PACS, and three other SCs used hybrid variants. Those using their own LSSarchive or hybrid also stored their exams with NLST PIDs rather than with medical-center IDs.

An SC retrieved exams from its medical center's PACS using a DICOM-transmission protocol in either or both of two modes: "query/retrieve" or "push". In the push model, the user runs an application on a PACS workstation or the PACS console. The user queries the PACS database, selects the exam and directs the PACS to transmit that exam to the CSW laptop. In the query/retrieve model, the query is sent by the CSW to the PACS. When an exam list is returned, the user selects one exam, and the CSW application sends a request to the PACS to retrieve that specific exam. The models differ in the location of the application that sends the query (directly on the PACS or remotely from the laptop).

Figure 1a and b describe the de-identification process. Figure 1a provides an overview: (1) DICOM exams are transferred from PACS to the laptop's folder (2) *Original DICOM Images* and (3) the monthly Westat Matching-List is loaded to the (4) *Westat Matching-List* folder. (5) These two folders are loaded to the de-identification application (6), and the user de-identifies exams as detailed in Figure 1b. (7) As each exam is de-identified, it is saved in the folder (8) *De-identified DICOM Images*. (9) De-identified exams may be copied to an external hard drive (XHD) or DVD and shipped to the CTIL; or they may be

CLARK ET AL.

transmitted via a virtual private network (VPN) through the Internet. Figure 1b gives data-level details of the de-identification process. (1) The deidentification graphical user interface presents a list of exams to the user, each line representing a single exam. (2) The information displayed for any exam is obtained from its original DICOM header (lower left). When the user selects an exam to be de-identified, the application searches the (3) Westat Matching-List until it finds a line where (4) DOB, Gender, and Exam Date match those of the selected exam's DICOM header. (5) If a match is made, then a new de-identified DICOM header is created (lower right). In this new header, the NLST-PID replaces the SC's medical



Fig 1. a De-identification overview. b. De-identification details.

center ID, and the screening year {T0, T1, T2} is put into a Comment field; also, in the new header, the patient's name is replaced with a generic "PATIENT^NAME" string, the DOB and Gender are blanked, the exam date is replaced by a fixed "19990101" string, and the accession number is blanked. Other DICOM fields containing protected health information are likewise blanked though not shown here. For the unlikely, though possible, case of two or more persons of the same gender with the same birth date being scanned on the same day (in the Westat list, the lines appeared identical except for the PID), the application recognized the multiplicity and prompted the user to choose the correct PID for the deidentification. An unmatched exam implied any of a number of things: (1) the Westat Matching-List that included this exam had not yet been received and loaded to the laptop; (2) the exam was not a valid screening exam; (3) the exam date, gender, or DOB in either the selected exam or the Westat list was wrong and in need of remedy. De-identified exams were saved in a temporary "export" folder on the laptop's hard drive.

Delivery Methods. Contents of the "export" folder could be copied, with other, similarly de-identified exams, to the XHD or to DVDs for shipment to the CTIL; or exams could be transmitted from the laptop over the Internet through a virtual private network. Choices and timing were dictated by SC workflow and shipping expenses. Any combination of methods could be used at any time. If the CTIL received an external hard drive, a blank drive was delivered to the originating SC the following day.

*Strategies for Collection.* Given that SCs were asked to collect, de-identify, and deliver thousands of CT exams, an unforeseen task assigned well after the launch of NLST enrollment and initial screenings, the SCs were given wide latitude in planning local workflow. SCs could deliver their CT screens in any order suiting local workflow, and no periodic quotas were imposed, though all exams were due by the end of 2006. For example, a site might submit all of its T0 exams, then its T1 exams, and finally its T2 exams while another site might submit all exams for participant-1, then for participant-2, et cetera. Still another site might submit most recent exams first, gradually working backward to the first screening exam.

#### **CTIL** Activity

Workforce. A radiologist directed the CTIL effort as principal investigator while the ERL director and a project manager with multicenter clinical-trial imaging de-identification experience provided general CTIL oversight. A general manager supervised day-to-day operations and dialogued with SCs regarding problems and issues that arose. A software specialist designed and maintained a database and a web-based application that accessed this database, permitting two image librarians to effectively direct their daily workflow. A systems manager and a network manager kept machines and communications working smoothly. Part-time image viewers assisted the librarians with the visual inspection of images for QA purposes. The CTIL principal investigator instructed librarians and image viewers in the techniques of inspecting lung CT images in the context of the NLST screening protocol.

Workflow. Exams arriving at the CTIL were digitally analyzed to make sure that each exam (a) corresponded to an actual, unique NLST screening CT exam, (b) contained an image-series meeting NLST-protocol specifications, and (c) contained no protected health information (PHI) in its DICOM headers. Exams were also visually inspected to make sure they (a) contained images of the lungs, void of significant artifacts, (b) contained no PHI, and (c) contained no imagemarkup annotations. Exceptions required CTILradiologist review, dialog between CTIL and SCs, or both. Exams passing these digital and visual-QA steps were archived into the CTIL. The archive hardware/software consists of mirrored EMC (Hopkinton, MA, USA) 8-TByte Centeras with a Merge Healthcare (Milwaukee, WI, USA) FUSION Server front-end.

To accomplish this quality-assurance progression, a CTIL management PostgreSQL (Wolfville, Nova Scotia, Canada) database tracked the movement of exams through four steps, detailed below: *received, input, prelim-QA*, and *visual-QA*.

 Received exams were logged, and a table-ofcontents spreadsheet containing NLST PIDs, number of images, and the DICOM unique identifiers associated for each delivered batch of exams was created and emailed to the originating SC for cross-check.

- 2. Exams with atypically large or small numbers of images were withheld from subsequent processing steps pending further evaluation. All other exams were *input* to a holding area, awaiting *prelim-QA*.
- 3. During *prelim-QA*, a software-script automatically checked exam DICOM headers for the inclusion of proper LSS scanning protocoldetermined reconstruction filters with correct slice thickness and spacing. Duplicate exams were flagged for resolution. The *prelim-QA* also checked DICOM headers for the inclusion of information likely to contain PHI; suspicious exams were parked for supervisory investigation. Acceptable exams, matching a Westat-supplied list of exams-to-expect (a cumulative list, updated monthly), moved onto *visual-QA*.
- 4. During visual-QA, exams were visually inspected for adequate exam and image quality using a Merge Fusion Server image viewer. Even though exams had already been deemed diagnostically acceptable by interpreting radiologists at the SCs, it was important to ensure that the images had not been compromised during retrieval from local PACS, during de-identification, or during transmission to the CTIL. A radiologist trained image librarians and image viewers to visually inspect CTIL images to confirm all images for each exam had been delivered, that they were actually NLST-protocol lung images with full lung coverage, that the images had no annotations labeling or analyzing pathology, that they contained no unusual image sequences, and that they were free of PHI. Failing exams were detained until their issues were resolved, with or without SC dialog and/or CTIL-radiologist assistance. In those cases where one or more images were missing or corrupt, partially or completely, the CTIL database noted the exam as "problematic" and a comment field in the database noted the image number(s). Corrupted images were rare but did occur; when encountered, the originating SC was notified and asked to check their image source (and re-send the exam if the originals appeared uncorrupted). Likewise, if an exam was missing more than a few images, the SC was asked to confirm no more images were available or to re-send the exam if more images were found. For an exam with a majority of images missing, which the SC could not recover, the

exam was considered "unavailable". Exam quality and image quality will be important to CTIL consumers, be they clinical researchers or nodule-detection algorithm developers.

These four steps typically spanned three days because exams were advanced through stages at night, in batch fashion, when librarians and viewers were not present and the database could be properly updated. Total elapsed times depended upon the number and sequence of exam arrivals, backlog of exams to input and run through QAprocessing, disk resources, the availability of viewers, and problems encountered. Exams passing all four steps were declared *archived*. Machine and personnel infrastructure permitted multiple, parallel processing across the four steps.

### Coordination Among Screening Centers, CTIL, Westat, and NCI

Periodic communications among four groups (SCs, Westat, NCI, and the CTIL) helped keep the project on task. The SCs were required to provide progress reports to NCI (monthly for the first 6 months, quarterly thereafter), and the CTIL provided weekly progress reports to Westat and NCI. Monthly conference calls of the LSS QA Working Group and the SC Coordinators allowed personnel from the four groups to discuss salient issues in a timely manner. Twice-yearly steering committee meetings, that included members of the four groups and LSS radiologists, served as fora for the presentation of formal status reports and the opportunity to discuss pressing issues.

#### RESULTS

*SC Strategies for Collection.* A variety of strategies ensued until the backlog of screens already obtained by the SCs had been delivered to the CTIL. Four SCs collected all T0 exams, then all T1 exams, then all T2 exams. Six SCs collected all three screens for participant #1, then those for #2, et cetera. Two SCs chose a combination. After the backlog had been discharged, exams were more typically accumulated, at the SCs, as their participants were screened; exams were then delivered in quantities and frequencies related to shipping expense and effort. *Method of Delivery*. Nearly all CT exams were delivered on XHDs (94.0%) or DVDs (5.9%) while very few were transmitted by virtual private network over the Internet (0.1%). Most SCs chose to submit the bulk of their exams on XHDs, while one SC chose to submit its exams solely on DVDs. CT exams on an XHD containing very few exams were typically copied to another XHD from the SC, and the former was then formatted and made available as a future swap or returned to the SC for additional exams.

*SC CTIL-related Workforce.* Most SCs budgeted the work over a 3-year period, though one chose 1 year and another chose 2. For multiple-year budgets, six SCs allocated the same number of FTEs each year while five SCs projected diminishing numbers of FTEs. The average (±stdev) number of SC FTEs per year per 1,000 exams per year was 0.41 (±0.38), median 0.27, range 0.14–1.41.

*Delivery Patterns*. Figure 2 shows delivery patterns for all SCs, by month, from January, 2005. Points in the lower half are typically DVD or Internet deliveries while those in the upper half are more likely external hard drive deliveries. Some SCs chose to deliver large numbers infrequently while others chose smaller numbers more frequently.

*Cumulative Progress.* Initial deliveries of deidentified exams began in January, 2005. Time spent hiring and training image librarians as well as fine-tuning workflow operations prevented actual archiving until May, 2005. The presence of image-embedded PHI in some exams required modification of de-identification software to de-



Fig 2. Delivery patterns by screening center (A-L) and month. Gross view of the variability in shipment frequency and numbers of exams per shipment from screening centers. Observations above the 100 mark were typically shipments on external hard drives, those below on CDs or DVDs.

tect such exams prior to their delivery to the CTIL. By Fall 2005, it was apparent that an additional workforce of part-time image viewers would be needed. From December, 2005 through March, 2007, six part-time image viewers were hired for varying hours/week and duration of months. The two image librarians and six part-time image viewers averaged (stdev) 5,964 ( $\pm 2,859$ ) exams viewed, range: 2,383–11,864. The part-time image viewers viewed 36,096 exams while working variable numbers of hours (average 535; stdev 264; range 193–773) and viewing varying numbers of image exams (6,016; 3,376; 2,381–12,015), translating to FTE/1,000 exams (0.05; 0.02; 0.02–0.08).

Figure 3 shows cumulative numbers of exams by stage: received, prelim-QA'd, visual-QA'd, and archived. The number received includes duplicate exams and resubmissions (replacing problematic priors) tendered by SCs. The gap in exams received versus exams archived was traced to the need for additional viewers as well as the need to develop database-management tools of greater complexity than originally anticipated, a longer average time to fully process the exams than originally anticipated (requiring supplemental image viewers), and additional unforeseen examspecific problems that arose after initial collection design. For example, the discovery of imageembedded PHI during visual-QA led to the installation of more robust de-identification software at both the SCs and CTIL prelim-QA

Exams (000's) Cumulative Exams by Processing Stage



Fig 3. Cumulative exams by processing stage. Total received (49,750) included duplicate and problematic exams not further processed; total archived (48,547) included those input that passed prelim-QA and visual-QA. Most activity completed early 2007 as seen in the apparent convergence of stage-lines in *the chart's upper right*. The final year involved extraordinary effort to recover every available exam not yet submitted or previously submitted with problems.

checkpoints; the software was then run on all exams at the CTIL awaiting processing. By the end of December, 2006 (the original target completion date), only 37,798 (78% of 48,547 expected) had been delivered. Remaining exams were delivered in 2007, and archiving was completed in February, 2008. A good part of the final months' effort was spent resolving outstanding issues with problematic exams and verifying with SCs the specific exams that were unavailable (lost, corrupt, compressed). These resolutions were delayed, at times, because SCs had not budgeted for this reconciliation period, and personnel were not always readily available.

*Number of Archived CT Exams.* Of the maximum number of possible CT exams (51,927 or three exams from 17,309 participants), performed screens numbered 48,723 (94%). Screens not performed were due to participant withdrawal (voluntary, death, required by NLST protocol), but the details of participant withdrawals remain unknown to CTIL personnel. Of the performed screens, only 176 (0.36%) were unavailable (lost, corrupt, compressed) from the SCs during the CTIL collection period, leaving 48,547 (99.64% of 48,723 screens performed) actually delivered and archived.

Number of Exams and Images per Exam, by SC. SCs enrolled varying numbers of participants. Figure 4 shows the distribution of CT exams by SC, both the potential maximum number (three screens from every participant) and the actual number received and archived. The number of image slices per exam varied for many reasons, among them: participant size, protocol applied (reconstructed slice thickness and interval), and number of separately reconstructed series per exam. Some SCs saved only a required single protocolspecified image series, while other SCs reconstructed and saved multiple series. Figure 4 also shows the average number of slices per exam by SC and the variation within each SC. For each SC, the lighter gray bar is the actual number of exams archived (scaled to the left ordinate); adding to that, a darker gray cap yields the maximum number of exams had all participants received three screens (i.e., no drop-outs). A circled center of a two-stdev error bar (scaled to right ordinate) is the average number of slices/exam from that SC. Overall, the average number of slices per exam was 257.



Fig 4. Exams and average slices/exam, by screening center.

PHI Detections. The potential for transmission of PHI was anticipated, but the locations in which it actually appeared were unexpected. Despite successful de-identification of the DICOM elements, we encountered exam dates in patient-protocol text-only image series, demographics in scout images, and radiology reports in secondary-capture image series. This required immediate notification of the originating SCs, Westat, and NCI, and simultaneous suspension of further collections from all SCs. Prior to resuming collection, the de-identification software was patched to detect and remove these kinds of image series. The upgraded software was then delivered to all SCs, the SCs ran the software and provided evidence of successful implementation, and collection was resumed. At the CTIL, all unprocessed exams, both those which had and had not yet arrived, were subjected to the same checks addressed by the software patches delivered to the SCs. Though twice applied (at the SCs and at CTIL), the checks made by these patches required but a few seconds per exam.

*Problems*. The vast majority of exams were delivered and archived without incident. However, multiple unexpected problems were encountered. Some could be solved in such a way as to prevent recurrence while others could only be solved semi-automatically; for example, an image series that contained multiple reconstructions needed to be divided into separate series, each containing images from a single reconstruction. Other issues required engagement of SC personnel; for example, a series lacking full lung coverage was accepted "as is" only after the SC confirmed it had delivered all of the images in its possession. As collection began, exams with problems detected by QA processes, automatic and visual, were "parked" for resolution while non-problematic exams were advanced through the system. Parked exams were processed in the background, when other activities were slack and/or CTIL management was available for resolution analysis. Librarians paperdocumented such problems and filed them by SC and PID. Eventually, the CTIL management database was modified to record such problems, and this facilitated problem resolutions. The majority of problems encountered are found in Table 2.

*CTIL Workforce*. The principal investigator (5%) FTE), laboratory director (5%), and project manager (5%) provided overall direction. A general-manager data administrator (50%) and database manager (40%) supervised two QA image librarians (each 75%) and a variable workforce of six part-time image viewers, working 8–32 h/week, no more than three at a time, for various numbers of months from December, 2005 through March, 2007. A systems administrator (15%) and network administrator (10%) provided installations (hardware and software), upgrades, and troubleshooting. NCI (<5%) and Westat (<5%) representatives monitored progress and coordinated with SCs to ensure timely exam-delivery.

Hardware/Software Failures. Two failed laptop hard drives were promptly repaired/replaced under a warranty agreement, as was one laptop's DVD writer. Two of 60 XHDs failed. In one case, all but 50 of 1,504 exams were rescued with special salvage software (Undelete 4.0; Executive Software International, Burbank, CA); and the SC was asked to resend the remaining exams that were unrecoverable. In the other case, all exams had been copied from the XHD and queued for processing, so it was unnecessary to ask the SC to re-send. Later, when trying to verify a problem with one of the exams, it was discovered that this XHD was unreadable and in need of re-formatting. In the EMC Centera mirrored archive, two nodes failed; because of the built-in redundancy, no data were lost. In late 2006, we were plagued with bottlenecks in the Merge Healthcare interface to the archive; but this was remedied with additional disk storage.

#### DISCUSSION

While the 48,000 exams collected for this research archive may pale in comparison to even annual accruals of many medical centers' PACSs, many of which also accept exams transmitted from multiple sites, this effort differed in several aspects. In contrast to the collection of clinical CT exams by a medical center's PACS, this creation of a CTIL required: (a) retrieval of archived CT exams from multiple SCs that employed various storage systems; (b) on-site de-identification of exams by each

Problem	Resolution
Receipt and input	
Atypical image count (too few, too many)	Verify with SC before proceeding.
Duplicate exams	Delete duplicates; notify SC. One SC with significant duplicates re-examined and modified their workflow strategy. If differing numbers of images, resolve with SC.
Prelim-QA	
No protocol-approved reconstruction filter	Petition SC for series with protocol-approved filter. Proceed with available series if protocol-approved filter is unavailable.
Patient-protocol, scout-image, secondary-capture series	Series excluded from archive.
Visual-QA	
PHI in images (radiology reports, demographics in scout images, patient-protocol exam dates)	SC software upgrade to eliminate secondary-capture series; same checks made during CTIL prelim-QA. Visual inspection still required.
Radiologist annotations, lossy compression	Accept exam but flag. Flagged exams may be filtered at the time of check-out.
Duplicate images in series	Manually remove duplicate images.
Merged series (multiple reconstructions)	Run script to separate reconstructions. Retain only protocol-valid series.
Multiple acquisitions in a series but unique image numbers	Log problem; pass on.
Incomplete lung coverage	Accept exam if nearly-complete-coverage; for obviously incomplete coverage, petition SC to check for more complete exam.

Table 2. Major Problems Encountered in CTIL Exams

individual SC; and (c) delivery of copies of all deidentified exams to the CTIL. And, unlike large clinical PACS collections, the CTIL and its associated participant demographic and medical histories will be made publicly available to lung-disease researchers and software developers once NLST follow-up data have been collected.

Many elements contributed to the successful collection of 48,000 CT exams: user-friendly software designed for efficient batch collection and deidentification of exams by SC personnel from their local archives; avoiding constraints on the order in which the serial individual-participant exams were retrieved and de-identified at the SCs, delivered to the CTIL, and processed and archived; largely reliable equipment; effective database-management software directing detailed QA workflow; and a cooperative spirit among SCs and CTIL teams. The groundwork laid by the SC and ERL/MIR personnel already involved in ongoing image QA procedures for the LSS facilitated the success of the CTIL collection process. SC personnel were already familiar with de-identifying and transferring images. CTIL staff had acquired knowledge of the technical parameters needed for the CTIL database that were located in the DICOM elements of image headers from different scanner models. As well, the OA experience enabled CTIL managers to understand the scope and detail of the hardware infrastructure needed and chosen for the collection project.

The quality of the exams in the CTIL can only be assessed in light of the QA procedures in place to clear exams for inclusion. Though timeconsuming and rigorous, those procedures applied to CTIL exams lend credence to the high quality of the library; as such, both automated analysis of DICOM image headers and visual inspection of images are crucial to the formation of any reputable image collection. While only a small fraction of CTIL exams were problematic, identifying and understanding problems retarded processing and detracted from optimal throughput. Recurring problems were resolved more rapidly as experience was gained. Many problematic exams were set aside, pending later resolution, but eventually required handling on an individual basis. Other problems required processing stoppages pending the design of software to query all received, unprocessed exams for similar problems (such as embedded PHI or screen-captured images), and implementation of preventive mechanisms at the SCs for uncollected exams. Problems were noted on work-lists of exams being processed, but the time-delays involved in servicing problems were not. Database-tracking of these problems and their delay times would have been preferable, but extensive exception accounting had not been budgeted, nor had we anticipated the number and variety of problems encountered.

The software developed for on-site, screeningcenter de-identification<sup>7</sup> has proven effective for CTIL collection. The same software tailored for two other trials, the Polycystic Kidney Disease Treatment Network<sup>8</sup> and the Silent Infarct Transfusion Study<sup>9</sup>, has met with equal success. Given that there was little time to design and test a more robust software suite dedicated specifically to CTIL image collection and de-identification, the software proved remarkably workable, though perhaps not optimal. Feedback suggestions from site users in all three trials should prove invaluable in the forging of the next, more robust version of this software that has now established an image-based clinical-trials track record.

The increasing emphasis on privacy concerns in the digital age requires close attention to management of PHI in clinical trials. Our experience reinforces the importance of implementing effective quality-control measures in the inter-institutional transfer of medical images, to insure compliance with the Health Insurance Portability and Accountability Act and IRB guidelines. Although laborintensive, the visual inspection of all images delivered to the CTIL was a critical component of this quality control, as the PHI encountered, in rare instances, was not present in DICOM headers, and thus had escaped the automated QA process. This step of visual inspection provided the ability to suspend any further collection and processing of images in a timely manner, and implement preventive software patches prior to resuming collection.

While NLST screening has been completed, the participant follow-up period extends into 2009. Earlier release to external investigators of large numbers of images tied to clinical information potentially could jeopardize the integrity of the NLST and interfere with its primary aims. Nonetheless, it may be possible to provide subsets of images to external investigators, with minimally necessary clinical information, depending upon the approval of specific projects by NCI and the trial's Data and Safety Monitoring Board. Such early releases provide an opportunity for the CTIL administrators to understand how large image sets are best provisioned to external investigators and how that provision should be managed.

There have been three formal requests for CTIL check-out of exams to NLST investigators. The requests illustrate the ways in which the CTIL might be utilized: one request was for a reader-variability study, one for a CAD development project, and one for comparison of emphysema in two different groups of NLST participants. While these requests were small in number (100, 100, 570 exams), these helped develop internal workflow and to test release mechanisms to investigators. Images were transmitted to one investigator via electronic network and to the others on portable media. The mechanism by which exams are supplied to fulfill future requests will depend on the size of the request and preference of the investigator.

Exam retrieval from the CTIL is a two-step process. (1) A query is posed to the CTIL management database using CTIL accession numbers selected by Westat, or any combination of criteria saved in the database to return CTIL accession numbers for exams meeting the criteria. Such criteria might be exam year, visit number, number of images, or any of the DICOM tags listed in Appendix B. (2) The file-list of accession numbers is passed to the PACS-like Merge Fusion Server that returns the image exams or requested series of images within those exams. External investigators pose their queries, based on demographic and/or clinical criteria, to Westat, keepers of the LSS clinical (non-image) database. For example, an investigator might ask for 180-image exam sets {T0, T1, T2} of females, aged 60–70, with lung nodules of size Xdetected at T2; Westat would then provide the CTIL with the corresponding CTIL accession numbers; the CTIL would pull the exams, as in (2) above, and ship them to the requesting investigator. These internal and external mechanisms are, thus, very flexible in tailoring queries to retrieve target images. Exact web-based mechanisms for making CTIL exams generally available are yet to be forged, but efforts are underway to use the National Cancer Imaging Archive (NCIA)<sup>10</sup>.

The NCIA aims to produce a publicly available searchable national repository integrating in vivo cancer images with clinical and genomic data in order to (a) improve the efficiency and reproducibility of cancer detection, diagnosis, and lesion classification; (b) accelerate diagnostic imaging decisions; and (c) effect quantitative and objective assessment of therapeutic response, eventually enabling the development of imaging resources, including validation in medical image processing<sup>11</sup>, leading to improved clinical decision support<sup>10</sup>. This NCIA repository already hosts two lung-disease related collections: (1) the Reference Image Database to Evaluate Response (RIDER) to therapy in lung cancer<sup>12</sup>, and (2) a database of spiral CT lung images assembled by the Lung Imaging Database Consortium (LIDC) to develop consensus guidelines for a spiral CT resource<sup>13</sup>. The CTIL and the images collected via the ACRIN component of the NLST offer their own uniqueness and significant volume, and could potentially be combined with these NCIA collections.

In the larger picture, image repositories such as the NCIA, hosting a variety of image collections, are likely to play significant roles in the cancer Biomedical Informatics Grid (caBIG)<sup>14</sup>, the NCI initiative to accelerate research discoveries and improve patient outcomes by linking researchers, physicians, and patients throughout the cancer community. As part of caBIG, not only are images available but also, perhaps, associated medical histories, histological findings, and genomic and proteomic data; and, over time, image knowledge-bases will grow with continued curation as more is learned about the images and their pathological content better understood. caBIG represents NCI's biomedical informatics efforts, modeled as a federated "grid" of interoperable research information systems (caGrid)<sup>15</sup>, to transform cancer research into a more collaborative and effective endeavor. The ERL/MIR team participated in a prototype development project exploring how the NCIA and caBIG might be leveraged to manage pathology data arising from the NLST program. This experience has led to the establishment of an NCIA instance in ERL/MIR. Perhaps in the future, this NCIA instance might be used to make subsets of the CTIL (and, eventually, associated clinical data) available to the cancer research community. The ability to view lungcancer nodules in CT images (from serial screens) side-by-side with digitized pathology slides and proteomic analyses may stimulate lung-disease researchers to think in different, more global ways, and facilitate more open collaborations among chest radiologists, pathologists, and medical geneticists.

*Strengths and Limitations.* The chief strengths of the CTIL are: (1) it contains a large volume of exams; (2) its images are from a very specific age-

range population meeting narrowly-defined inclusion criteria of heavy-history smoking without lung cancer; (3) there are serial scans for a large percentage of these participants; (4) the participants were scanned according to a defined protocol that allowed the use of multiple scanner vendors and models at different sites; (5) the images were uniformly de-identified on the same laptop models with identical software at all sites; (6) images arriving at the CTIL were scrutinized with strict quality-control procedures applied to DICOM headers and visual inspection of images; (7) clinical and demographic data associated with the images was collected and managed with extreme attention to detail independent of the CTIL; (8) images are available to investigators on a per request basis tied to a research plan that is independent of NLST end-point determinations. The chief limitation of the CTIL is its temporary unavailability to public access because the NLST has not completed clinical follow-up to determine those end points and will not until at least late 2009. In addition, although most of the scanner models used to acquire these images are still in use, they no longer represent current stateof-the art equipment.

Lessons learned in the design, implementation, and execution of this CTIL collection may be helpful to planners of future large-scale multicenter imagingbased clinical trials. (1) Images may be centrally archived at a site distinct from a data coordinating center, but coordination and communication are required to ensure that the expected imaging exams arrive at the image archive. Because study personnel at sites contributing images may have limited imagetechnology expertise, establishing congenial rapport, providing detailed written instructions with updates as needed, and being readily available for problemsolving are crucial to rapid implementation and overall success. (2) Image collection and processing procedures must be flexible to accommodate the unanticipated problems that will arise. Solutions must be forged promptly and dispatched to all sites in order to avert problem repeats. Anticipating likely problems and building solid quality-control procedures will likewise pay time-saving dividends through the trial's duration. (3) Periodic conference calls and steering committee meetings are helpful in discussing common problems and issues as well as publicly comparing and encouraging site-by-site progress. (4) Within the image-archiving site, weekly progress charts encourage a sustained high level of effort; in turn, periodic reports to the data coordinating center encourage image-archive site management to keep charges on task and affirm their steady contributions. (5) Helpful software tools for exam processing and database management are essential, though they may require custom design to match specialized needs. For example, electronic worksheets for librarian tasks were essential in keeping this project on task. As well, allowing our librarians to suggest better methods for workflow-monitoring software and then participate in its testing and retesting improved internal processing. (6) Tight quality control is essential and involves scrutiny of DICOM image headers and the images themselves. (7) Having multiple individuals process incoming exams keeps the project going when one is unavailable. Together, they complement one another and discover better, more efficient ways of doing things. Their value is enhanced when their time is shared across multiple trials and insights from one trial may aid another. (8) A mechanism for electronically capturing and quantifying problems is helpful for tracking the changes in problem-type frequency and preventing their recurrence.

#### CONCLUSION

The approach to large-scale, multicenter CT image collection described here may serve as a useful model, while the experience reported should be valuable for refining the planning and execution of future equivalent endeavors.

#### ACKNOWLEDGEMENTS

This research was supported by contracts from the Division of Cancer Prevention, National Cancer Institute (NCI), NIH, DHHS. The authors thank Drs. Christine Berg, LSS-NLST Project Officer, and John Gohagan, former LSS-NLST Project Officer, Division of Cancer Prevention, National Cancer Institute; the Screening Center (Appendix A) investigators, coordinators, and staff of the National Lung Screening Trial (NLST); Mr. Tom Riley and staff, Information Management Services, Inc., and Ms. Brenda Brewer and staff, Westat, Inc, for their support and assistance. The Westat LSS component is supported by NCI Contract NO1-CN-25476. We thank Drs. Richard Fagerstrom and Timothy Church for their reviews of a preliminary version of the manuscript. We also thank our image viewers, without whom the CTIL would have been seriously delayed: Angelica Cosas, Patricia Rueweler, Rochelle Williams, Dr. Sooah Kim, Dr. Miyoung Kim, and Dr. Yuting Liang. The CTIL gratefully acknowledges Merge Healthcare's generous contribution of the FUSION Server and its continued support under their research

agreement with the Mallinckrodt Institute of Radiology. Most importantly, we acknowledge the LSS participants for their contributions to making this study possible.

#### APPENDIX A

#### Table 3. Lung Screening Study Screening Center and *Satellites* with their National Cancer Institute Contract Numbers

Institution (City, State)	NCI contract
Georgetown University (Washington, DC)	NO1-CN-25522
Henry Ford Health System (Detroit, MI)	NO1-CN-25512
Marshfield Medical Research and	NO1-CN-25518
Education Foundation (Marshfield, WI)	
Pacific Health Research Institute	NO1-CN-25515
(Honolulu, HI)	
University of Alabama at Birmingham	NO1-CN-75022
(Birmingham, AL)	
University of Colorado (Denver, CO)	NO1-CN-25514
University of Minnesota (Minneapolis, MN);	NO1-CN-25513
Center for Diagnostic Imaging	
(Indianapolis, IN)	
University of Pittsburgh (Pittsburgh, PA)	NO1-CN-25511
University of Utah (Salt Lake City, UT);	NO1-CN-25524
St. Luke's Regional Medical Center	
(Boise, ID)	
Washington University (Saint Louis, MO)	NO1-CN-25516

#### APPENDIX B: CT PARAMETERS

Researchers using the CTIL will require access to the attributes that are found in the DICOM images. These describe the techniques used for each acquisition as well as the equipment that is used. These attributes, from each image series of each CTIL exam, have been forwarded to Westat in order that Westat might readily respond to a researcher's formal request for CTIL images by determining which exams will satisfy the researcher's needs; otherwise, each request involving these attributes would require the extra step of first searching the CTIL database.

Table 4 lists DICOM attributes for values of interest during the screening trial and for secondary analysis. Not all LSS scanners provide values for each of these attributes. In some cases, the attribute stored in the database is actually calculated from other data in the DICOM image. In other cases, the images do not contain sufficient information to calculate or infer those values. The value stored in the CTIL database for each attribute is dependent on the vendor. For example, a manufacturer may provide a value for table rotation but not a value for collimation; or, those values may be entirely absent (Table 4)

Note 1: in order to calculate 'effective mAs' (=mAs/ pitch), one needs to first determine mAs and pitch. Both of those required extensive reading of DICOM conformance statements and explanations from the manufacturers. Values needed to determine mAs and/ or pitch are sometimes stored in private attributes. Discussions with the manufacturers' engineers allowed us to interpret those private attributes and extract the

#### Table 4. DICOM Tags Stored in CTIL Database

Attribute	DICOM Tag	Comment
kVp	0018 0060	
Exposure (mAs Calculated)	0018 1152	Calculated as exposure time times x-ray tube current
		Most of the equipment did not store this directly. The Siemens
Exposure (mAs Direct)	0018 1152	equipment placed effective mAs in 0018 1152
Effective mAs (calculated)	0018 1152	See Note 1
		Only Siemens stores effective mAs directly. This is found
Effective mAs (direct)	0018 1152	in 0018 1152
		For some equipment, exposure time found in a private element
Exposure Time	0018 1150	and not in the DICOM standard element
X-ray Tube Current	0018 1151	
Table Feed per Rotation	0018 9310	(When available)
CT Pitch Factor (calculated)		See Note 2
CT Pitch Factor (direct)		See Note 2
		We automatically checked each series for uniformity of slice
Slice thickness (reconstruction thickness)	0018 0050	thickness in the images.
Pixel Spacing (reconstruction interval)	0028 0030	See Note 3
Reconstruction filter	0018 1210	
Manufacturer	0008 0070	Scanner manufacturer name
Manufacturer's Model Name	0008 1090	Scanner model
Software Version	0018 1020	Scanner software version at time of scan
Scanner ID		See Note 4

Note 2: values for pitch were never found directly in a public DICOM attribute. In the Philips equipment, the value for pitch was found in a private attribute. For GE scanners, a text string in a private attribute stored a coded value used to look up the pitch. The Siemens scanners did not record pitch; we performed an inverse calculation based on the effective mAs stored by the equipment and our determination of mAs from exposure and current.

The Toshiba CT devices store several parameters in private elements that must be combined to compute pitch. The Toshiba CT system creates a private attribute with a binary object that includes a wide range of data. Among other things, these data include the required CT pitch information but also PHI captured by the modality. Passing that attribute from the Toshiba system through the laptop to the CTIL would result in PHI disclosure; omitting that data would exclude the important CT pitch information. This problem was addressed by modifying the laptop software to interpret the Toshiba binary object, extract the required pitch information, and encode the pitch information in a separate private attribute. The original Toshiba private attribute could then be safely deleted. This is the only instance where the laptop de-identification software was modified to interpret data specific to a manufacturer or device model. All other manufacturer specific data (non PHI) were interpreted with custom software maintained at the CTIL.

Note 3: pixel spacing (reconstruction interval) was computed from Image Position (0020 0032) rather than just extracted from (0028 0030). This provided a backup mechanism to allow us to check for uniformity of slice location to make sure that slices were not missing. This is best done by examining Image Position (0020 0032) and not making assumptions about Instance Number (0020 0013).

Note 4: scanner ID was a scanner identifier copied from a cross-reference table of identifiers built from the DICOM attributes Institution (0008,0080), Station Name (0008, 1010), Manufacturer (0008,0070), Manufacturer's Model Name (0008, 1090), and Software Version (0018, 1020). The table had been built from information provided by LSS medical physicists for known scanners for which the physicists had routinely supervised calibrations as part of LSS quality assurance. If the attributes in the DICOM header of any CT exam delivered to the CTIL could not be matched in this table, the originating site was questioned because the exam seemed to have come from a scanner not being monitored for quality assurance. Such exceptions were fewer than ten; and most of these cases were caused by scanner software upgrades of which the CTIL had not been informed.

#### REFERENCES

1. Gohagan J, Marcus P, Fagerstrom R, Pinsky P, Kramer B, Prorok P, for The Lung Screening Study Research Group: Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph (The Lung Screening Study of the National Cancer Institute). Chest 126:114–121, 2004

2. Hillman BJ: Economic, legal, and ethical rationales for the ACRIN National Lung Screening Trial of CT screening for lung cancer. Acad Radiol 10:349–350, 2003

3. Gohagan JK, Prorok PC, Hayes RB, Kramer BS, and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Project Team: The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials 21(6 Suppl):251S– 272S, 2000

4. Moore SM, Gierada DS, Clark KW, Blaine GJ: Image quality assurance in the Prostate, Lung, Colon, and Ovarian (PLCO) Cancer Screening Trial Network of the National Lung Screening Trial. J Digit Imaging 18(3):242– 250, 2005

5. Clark KW, Gierada DS, Moore SM, et al: Creation of a CT image library for the Lung Screening Study of the National Lung Screening Trial. J Digit Imaging 20(1):23–31, 2007

6. Cagnon CH, Cody DD, McNitt-Gray MF, Seibert JA, Judy PF, Aberle DR: Description and implementation of a quality control program in an imaging-based clinical trial. Acad Radiol 13(11):1431–1441, 2006

7. Moore SM, Maffitt DR, Blaine GJ, Bae KT: A workstation acquisition node for multi-center imaging studies. SPIE Medical Imaging 2001, PACS and Integrated Medical Information Systems: Design and Evaluation 4323:271–277, 2001

8. Washington University in Saint Louis—Polycystic Kidney Disease Treatment Network (PKD-TN). Available at http:// www.pkd.wustl.edu/pkd-tn. Accessed 24 June 2008

9. Washington University in Saint Louis—Silent Infarct Transfusion(SIT) Study. Available at http://sitstudy.wustl.edu. Accessed 24 June 2008

10. National Cancer Institute—National Cancer Imaging Archive (NCIA). Available at http://ncia.nci.nih.gov. Accessed 24 June 2008

11. Jannin P, Krupinski E, Warfield S: Validation in medical image processing (guest editorial). IEEE Trans Med Imag 25 (11):1405–1409, 2006

12. National Cancer Institute—National Cancer Imaging Archive (NCIA). Reference Image Database Resource (RIDER) and Plans for a Public-Private Partnership (white paper). Available at http://ncia.nci.nih.gov/ncia/collections (under RIDER). Accessed 24 June 2008

13. Armato SG, McLennan G, McNitt-Gray MF, et al: Lung Image Database Consortium: developing a resource for the medical imaging research community. Radiology 232:739–748, 2004

14. National Cancer Institute—Cancer Biomedical Informatics Grid (caBIG). Available at https://cabig.nci.nih.gov. Accessed 24 June 2008.

15. Saltz J, Oster S, Hastings S, et al: caGrid: design and implementation of the core architecture of the cancer Biomedical Informatics Grid. Bioinformatics 22(15):1910–1916, 2006