

Kenneth Clark¹, Bruce Vendt¹, Kirk Smith¹, John Freymann², Justin Kirby², Paul Koppel¹, Stephen Moore¹,

Stanley Phillips¹, David Maffitt¹, Michael Pringle¹, Lawrence Tarbox¹, Fred Prior¹

¹Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110

²Clinical Research Directorate / CMRP, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702

Abstract

- The National Institutes of Health (NIH) have placed significant emphasis on the sharing of research data to support secondary research and encourage investigators to publish their clinical and imaging data.
- Realizing unlinked distributed image collections were not conducive to large scale secondary research studies, the National Cancer Institute (NCI) created the open source National Biomedical Image Archive (NBIA)^[1] to facilitate centralized hosting of cancer related images.
- NCI leveraged the NBIA database, the Clinical Trials Processor (CTP)^[2], and image hosting expertise of Washington University to create **The Cancer Imaging Archive (TCIA)** - an open source, open access information resource to support research, development, and educational initiatives utilizing advanced medical imaging of cancer.
- The complex challenges of operating and maintaining an open-access image archive that meets the scientific needs of image consumers while maintaining high-availability, confidentiality, ease of use, and security are presented.

Methods

- Workflow.** Figure 1 and Table 1 illustrate how the TCIA manages image submission, de-identification, and curation; image provider and consumer support; and workflow process with a dedicated set of tools, technology, and skilled personnel. Figure 1 sequential circled-numbers define the image-collection steps (comments adjacent to numbers). Table 2 shows both the front-end detailed information gathered from the image provider and the submission-schedule checklist of detailed steps.
- Quality Control.** TCIA follows a rigorous image curation process to insure images are properly de-identified prior to public dissemination. A customized instance of CTP is deployed at submitting sites to de-identify DICOM standard elements in accordance with DICOM PS 3.15, Appendix E: Attribute Confidentiality Profiles.^[3] Each arriving image is visually inspected for burned-in protected health information (PHI), private DICOM elements are de-identified based on a TCIA knowledge-base of published DICOM conformance statements, and unique element values from DICOM standard and private elements are printed and reviewed for PHI by TCIA curators.
- Personnel.** The TCIA operations and maintenance team consists of experienced project managers, image quality-control curators, subject matter experts (cancer, imaging, DICOM, technology), systems and network administrators, software developers to add new features, a customer support center, and experienced leadership to set and monitor priorities and continually focus on the overall vision for TCIA.
- Documentation.** Standard Operating Procedures and checklists insure established procedures are followed. A dashboard provides information to senior management for decision making. A private project wiki space is used to manage image-collection status (subset, Figure 2). A public wiki space provides collection-specific details and helpful links for users.
- Data Consumption.** Investigators downloading images from the TCIA public server do so through the NBIA application.

Figure 2. Submission Status at a Glance

Update	Priority/Status	Owner	Collection	Site	Patients	Images
5/10/2012	2/active	WUSTL	QIN Lung Segmentation Challenge	Moffitt	100	7,975
5/29/2012	1/active	WUSTL	TCGA-LUAD	WASHU		
11/8/2011	Inactive	WUSTL	TCGA-GBM	Duke		
5/23/2012	1/active	WUSTL	TCGA-GBM	TJU	46	34,182
5/15/2012	1/active	WUSTL	TCGA-LGG	TJU	26	15,411
5/4/2012	3/active	WUSTL	QIN HeadNeck	Iowa	142	280,134
5/25/2012	3/active	WUSTL	QIN-Breast	Vanderbilt	15	16,646
4/10/2012	3/active	WUSTL	QIN-Brain	UPMC	1	8,694
	Inactive	CIP	NaF-Prostate	Wisconsin		
5/30/2012	Complete	WUSTL	HeadNeck Cetuximab	RTOG	96	15,199
5/10/2012	2/active	WUSTL	Phantom FDA	FDA	1	240,835

Figure 1. The Cancer Imaging Archive (TCIA) Image and Information Flow

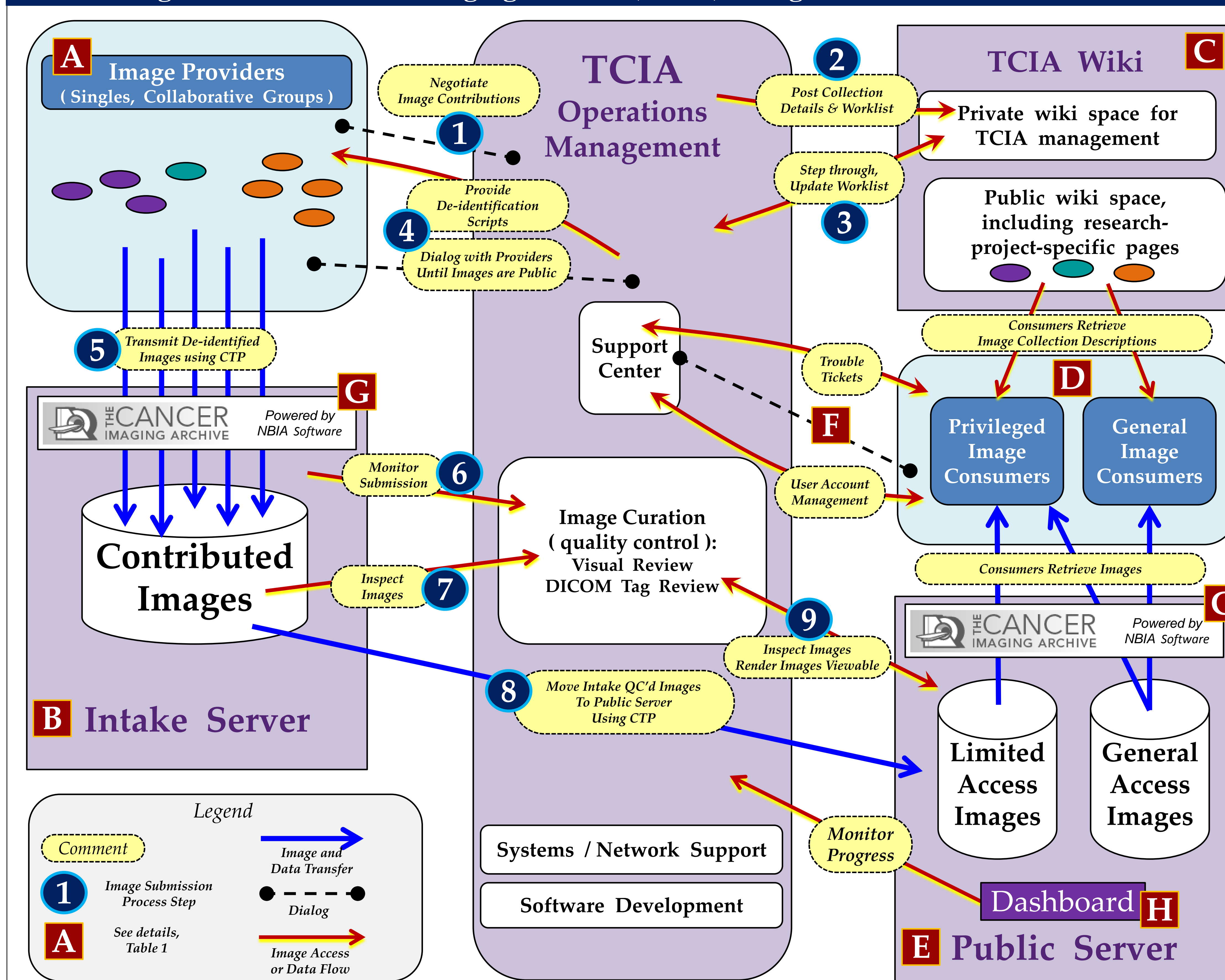


Table 1. Details Referenced in Figure 1

- A** Image submitters route images through a CTP pipeline that de-identifies images and queues them for transmission to the intake server. Images are de-identified according to DICOM PS 3.15, Appendix E: Attribute Confidentiality Profiles de-identification guidelines.
- B** The intake server accepts multiple concurrent submissions. These are curated (quality control review) before moving to the public server and made available to image consumers.
- C** The TCIA wiki is used for workflow management and supporting documentation. It also gives data contributors a platform to describe the scope and intent of their image collection and to provide metadata and/or ways for users to reach them for questions, etc. The wiki supports research groups by summarizing the work of participants, including the posting of conference abstracts and publications.
- D** Privileged image consumers have been granted special permissions to query/download collection-specific limited-access images, while "general" consumers cannot.
- E** The public server hosts thoroughly curated images available for download and the NBIA application for doing so. Downloads are noted in NBIA's MySQL database.
- F** The operations management Support Center's trouble-ticket system tracks image-consumers' issues and resolutions. Consumers reach the support center via email or an 800 telephone number. The Support Center also coordinates user account management.
- G** Image consumers view and download images via an application NBIA that also permits one to save lists of images that may be shared with colleagues. NBIA allows one to search for images by collection, anatomy, modality, scanner manufacturer, dates, and many modality-event specific criteria (CT slice thickness, MR pulse sequence, etc.)
- H** A dashboard reporting system provides TCIA management with current and historical consumer counts, collection image counts, and image downloads.

Table 2. Submission Status Details Template

Collection Information	Submission Schedule (date, initials, comment)
Primary Investigator Contact Information	Contacted Data Provider for Information
Technical Point of Contact Information	Create submitter's collection/site NBIA/UPT elements
Usage Restrictions of Data	Create CTP scripts/config files
Availability or Relevant Deadlines	Send CTP scripts/config files to Submitter
Batch Schedule for Submission	Image Test Transmission Complete
Data Transfer Mechanism	Submitter Begins Collection Transmission
Verify No Prior De-identification	Transmission Complete
Collection Access (Public or Limited)	Begin QC on Intake Server (Tag Analysis, Visual Review)
Modalities (MR, CT, PET, etc.)	Complete Intake Server QC
Patient ID Format	Submitter Asked to Review/Approve DICOM changes
Body Parts Examined	Collection Sign-off by Image Submitter
No. of Patients, Studies per Patient, Series per Study	Prepare CTP Script for Upload of Images to Public
Approximate Date Range of Image Studies	Upload Images to Public Server
Potential PHI in image pixels (Screen Save, Overlay)	Begin QC on Public Server
Private Tag Requirements	Complete Public Server QC
Accompanying Metadata (.xml, .pdf, .xls, etc.)	Test Retrieval with NBIA Download Manager
Would you like a wiki page for your collection	Collection complete; Update Wiki

Results (Year 1)

Figure 3 shows the number of contributed images and contributing sites. Figure 4 shows the various Support Center Tickets, predominantly for new user accounts; "general" tickets are often queries about specific collections. "All Other" includes "image" tickets, referring to images seemingly missing or unavailable; "new collection" tickets enquiring of new submissions; and "critical" tickets dealing with server and web site outages. Table 3 reports the total TCIA image count with public and limited-access percentages.

Figure 3. Number of Contributed Images and Contributing Sites (by Month)

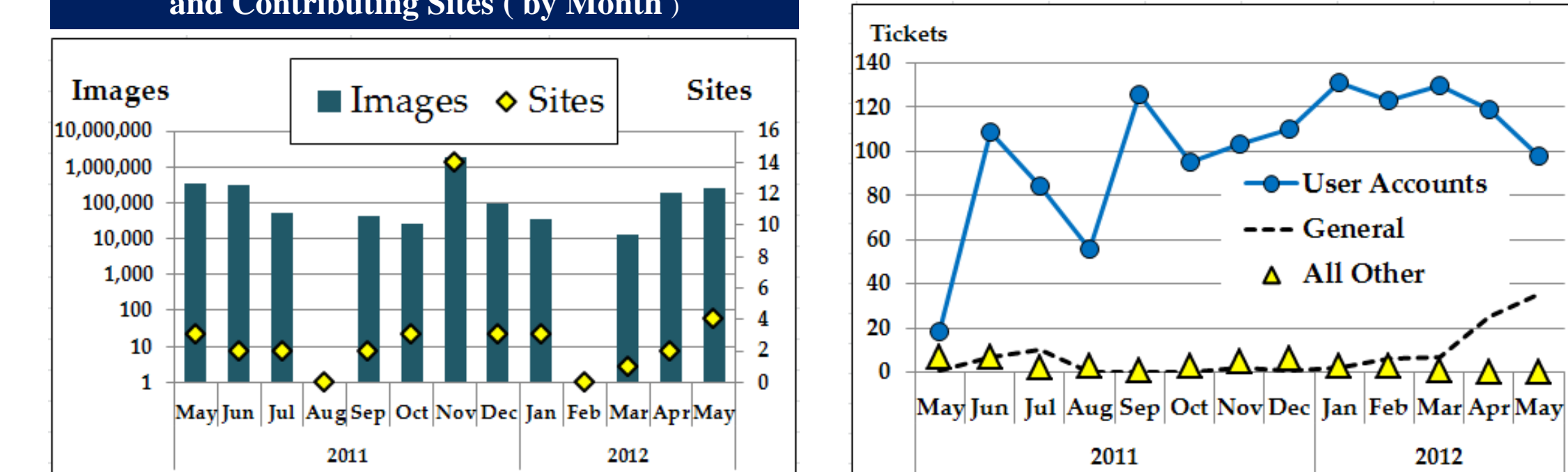


Table 3. TCIA Total Images

TCIA Total Image Count: 3,268,664 (Access: Public 96%, Limited 4%)

Conclusion

- The NCI has funded a managed resource to support a centralized collection of cancer imaging data.
- NCI's Cancer Imaging Program and Washington University have collaborated to build the infrastructure using NBIA open source software and to put in place processes and staff members who actively manage the archive and provide industrial level support to both end users and image providers.
- This resource helps those investigators who need to publish their collections as well as those who are looking for high quality data sets to further their research.

References

- TCIA User Guide (including description of the NBIA application). <https://wiki.cancerimagingarchive.net/display/Public/The+Cancer+Imaging+Archive+User%27s+Guide>
- CTP - The RSNA Clinical Trial Processor. http://mirwiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor
- DICOM Standard Part 15, Appendix E: Attribute Confidentiality Profiles, pp. 60-92, PS 3.15-2011 (de-identification guidelines) http://medical.nema.org/medical/dicom/2011/11_15pu.pdf

Acknowledgements

TCIA is funded by SUBCONTRACT 10XS220: SAIC-F (PI: Prior) Image Archive Hosting. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN26120080001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Contact: Ken Clark. (314) 362-1520. Email: clarkk@mir.wustl.edu