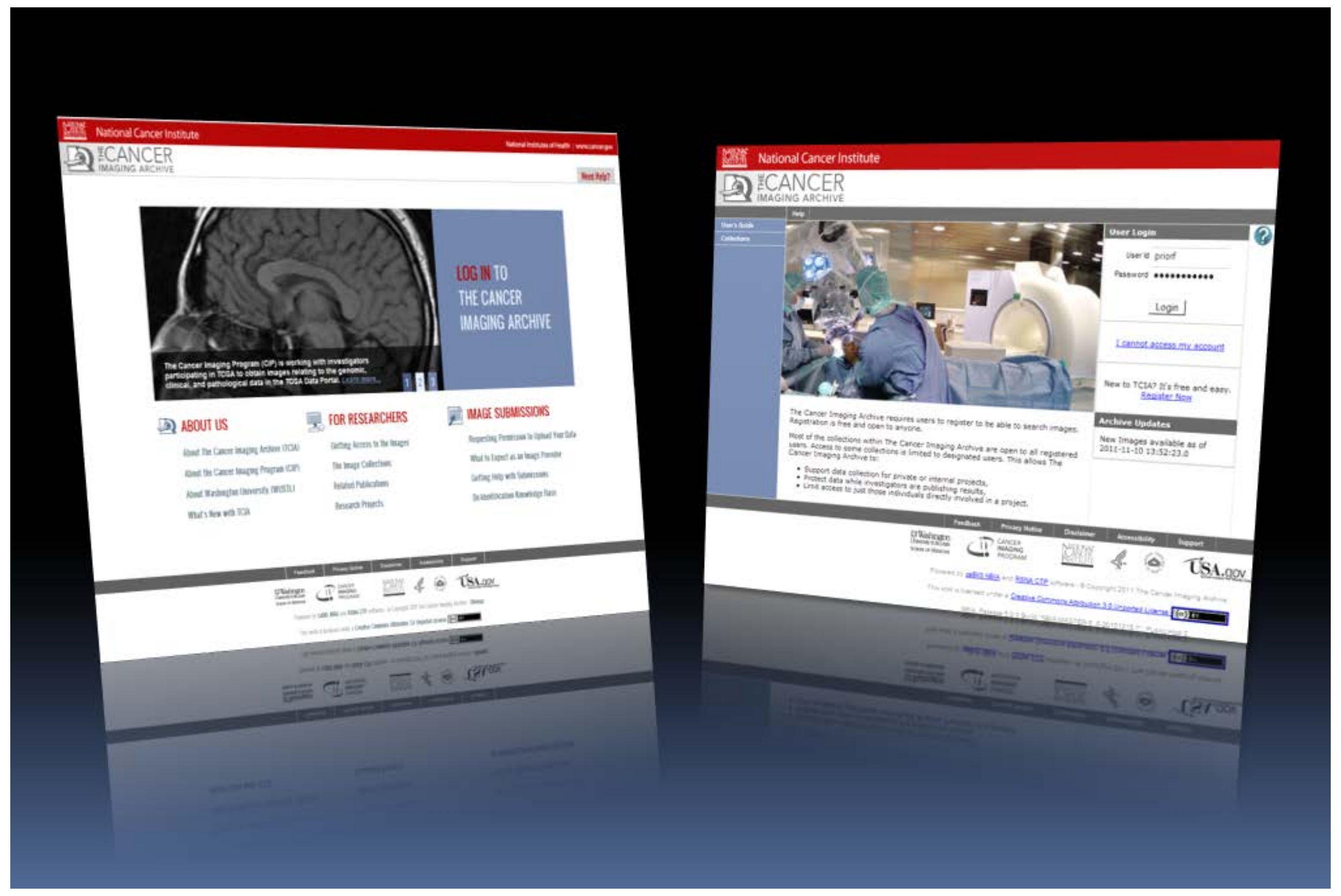
CANCER IMAGING ARCHIVE
a Repository of Advanced Imaging Information Correlated with TCGA Samples

Fred Prior[1], John Freymann[2], Bruce Vendt[1], Ken Clark[1], Justin Kirby[2], Lawrence Tarbox[1], Paul Koppel[1],
Stephen Moore[1], Mike Pringle[1], Stanley Phillips[1], David Maffitt[1], Carl Jaffe[3]
[1]Mallinckrodt Institute of Radiology, Washington University Medical School, St. Louis, MO 63110
[2]SAIC-Frederick, Inc. [3]Boston University School of Medicine

ABSTRACT: **Advanced medical imaging that is correlated with tissue samples and associated genomic analysis results provides a unique research resource. One of the goals of The Cancer Imaging Archive (TCIA) project is to collect, consistently de-identify, curate and make publicly available rich collections of imaging data. TCIA includes many of the imaging studies used to diagnose and characterize the solid tumors that were sampled for the Cancer Genome Atlas (TCGA) initiative. These data sets are identified consistently with TCGA to facilitate correlative studies. The open source National Biomedical Imaging Archive (NBIA) software package, developed through the Cancer Bioinformatics Grid (caBIG) initiative, was modified to improve download performance and enable hosting in a high availability, cloud computing environment. An IRB approved process for secure transport and consistent de-identification that preserves scientifically significant information in vendor proprietary data elements has been implemented. This process, based on the open source Clinical Trial Processor software package, includes automated analysis of vendor proprietary data elements and text fields for detection of protected health information as well and multi-level security protocols. Current image collections that are publicly available include 358,000 Magnetic Resonance images representing 285 studies of glioblastoma multiforme (GBM) that are cross-linked to the TCGA GBM data sets. Additional image collections including breast and lung cancer studies also linked to TCGA are underway. Currently over 1,725,000 images are available for download with 600,000 more in the incoming pipeline. Researchers from 51 countries have thus far downloaded over 10 million images. TCIA adds a new dimension to the Cancer Genome Atlas research program by enabling research on new quantitative and qualitative analyses that link imaging features to genomic signatures. TCIA also facilitates verification and validation of new computer aided analysis tools and imaging based candidate biomarkers.**

TCIA is a Large and Growing Archive Service Providing Images and Related Information for Use in Research



- **Cancer researchers can use this data to test new hypotheses and develop new analysis techniques to advance our scientific understanding of cancer.**
- **Engineers and developers can build new analysis tools and techniques using this data as test material for developing and validating algorithms.**
- **Professors can use it as a teaching tool for introducing students to medical imaging technology and cancer phenotypes.**

FIGURE 1: **TCIA may be accessed a via the Analytical Tools page from the TCGA web site: http://tcga-data.nci.nih.gov/tcga/tcgaAnalyticalTools.jsp or directly at: http://www.cancerimagingarchive.net .**

The landing page (left) provides linkages to information about the services provided by the TCIA and the available collections. The login page (right) provides access to the collections.

Image Collections Available from TCIA

- **TCIA is organized into Collections: typically groups of patients related by a research aim, common disease (e.g. lung cancer), or image modality (MRI, CT, etc).**
- **Additional information about the intended purpose for each Collection can be found on the TCIA Wiki**

TCGA	QIN	Other
GBM/ MDA	Breast/ Vanderbilt *	LIDC-IDRI/ CIP
GBM/Henry Ford	Head-Neck/ Iowa *	Breast-Diagnosis/ BU
GBM/ UCSF	Head-Neck/ UPMC *	Prostate-Diagnosis/ BU
GBM/ Duke	Phantom/ UW *	Prostate-MRI/ NCI-CCR*
GBM/ Emory	Phantom/ Maastro *	NaF/ NCI-CCR *
BRCA/ MSKCC	Brain/ UPMC *	CT Colonography/ CIP
BRCA/ Roswell Park	Prostate/ BWH *	RIDER / CIP
BRCA/ Mayo		QIBA CT
BRCA/ UPMC		
LUAD/ WUSTL		

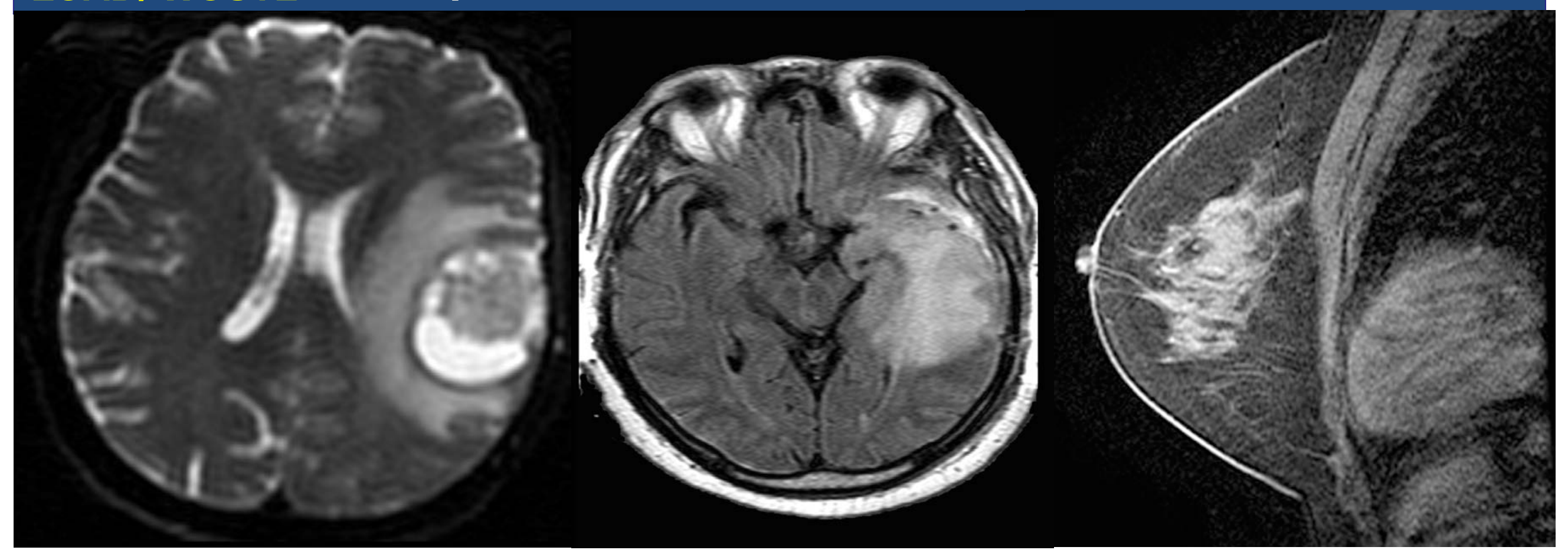


FIGURE 3: **Over 20 different institutions have already provided data to TCIA since it was announced in June of 2011. The table on the left indicates the collections that are currently AVAILABLE on TCIA (green) and IN PROCESS (yellow). The (*) indicates a private collection. The images on the right represent examples of MRI data sets linked to TCGA samples via a TCGA participant ID. The left image is a diffusion weighted MRI illustrating a GMB. The center image is a T2 FLAIR of a different GBM patient. The right –most image is a breast IMR from the BRCA study. Clinical, genetic, and pathology data linked to all TCGA image sets may be accessed via the TCGA data portal .**

FIGURE 4: **Image Curation is performed by an experienced team of experts who ensure all Protected Health Information (PHI) is removed from the data.**

- **All data is processed with the RSNA's Clinical Trials Processor (CTP) software before it leaves the sending institution using de-identification scripts which leverage DICOM Supplement 142 for clinical trials image de-identification [2].**
- **Automated tools scan DICOM image headers and vendor proprietary data elements to remove protected health information and retain scientifically valuable information.**

REFERENCES

1. Zinn PO, Majadan B, Sathyan P, Singh SK, Majumder S, et al. 2011 Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Multiforme. PLoS ONE 6(10): e25451. doi:10.1371/journal.pone.0025451

2. Freymann J, Kirby J, Perry J, Clunie D, Jaffe C. Image Data Sharing for Biomedical Research - Meeting HIPAA Requirements for De-identification. Journal of Digital Imaging. 2011:1-11.

How TCIA Enables TCGA Imaging Research Groups

- **CIP TCGA Radiology Initiative** — Driven by input from its scientific community, the Cancer Imaging Program (CIP) finds itself at the junction of two powerful scientific requisites; the need for cross-disciplinary research and inter-institutional data-sharing to speed scientific discovery and reduce redundancy, and the need to provide imaging phenotype data to augment large scale genomic analysis.
- **TCGA Breast Phenotype Research Group** — TCGA Breast Phenotype Research Group is part of the CIP TCGA Radiology Initiative. The group began as an ad hoc multi-institutional research team dedicated to discovering the value of applying controlled terminology to the MR imaging features of patients with breast cancer. This activity is currently in the early stage of development. MR images which correlate to the Breast Invasive Carcinoma (BRCA) data in the TCGA Data Portal are currently being gathered for submission to TCIA. In the mean time the group is beginning preliminary discussions around research methods and goals, as well as utilizing the existing Breast-Diagnosis collection as a training set their efforts.
- **TCGA Glioma Phenotype Research Group**— TCGA Glioma Phenotype Research Group is part of the CIP TCGA Radiology Initiative [1]. The group began as an ad hoc multi-institutional research team dedicated to discovering the value of applying controlled terminology to the MR imaging features of patients with gliomas. Research trials that incorporate imaging present unique challenges due to nonstandard use of terminologies, absence of uniform data collection and validation. These obstacles traditionally limit the impact of imaging as an effective biomarker in oncology. The purpose of this project was to assess reliability of tools and terminology developed by the Cancer Bioinformatics Grid (caBIG) initiative when performing a multi-reader simultaneous assessment of glioblastoma MR imaging features.

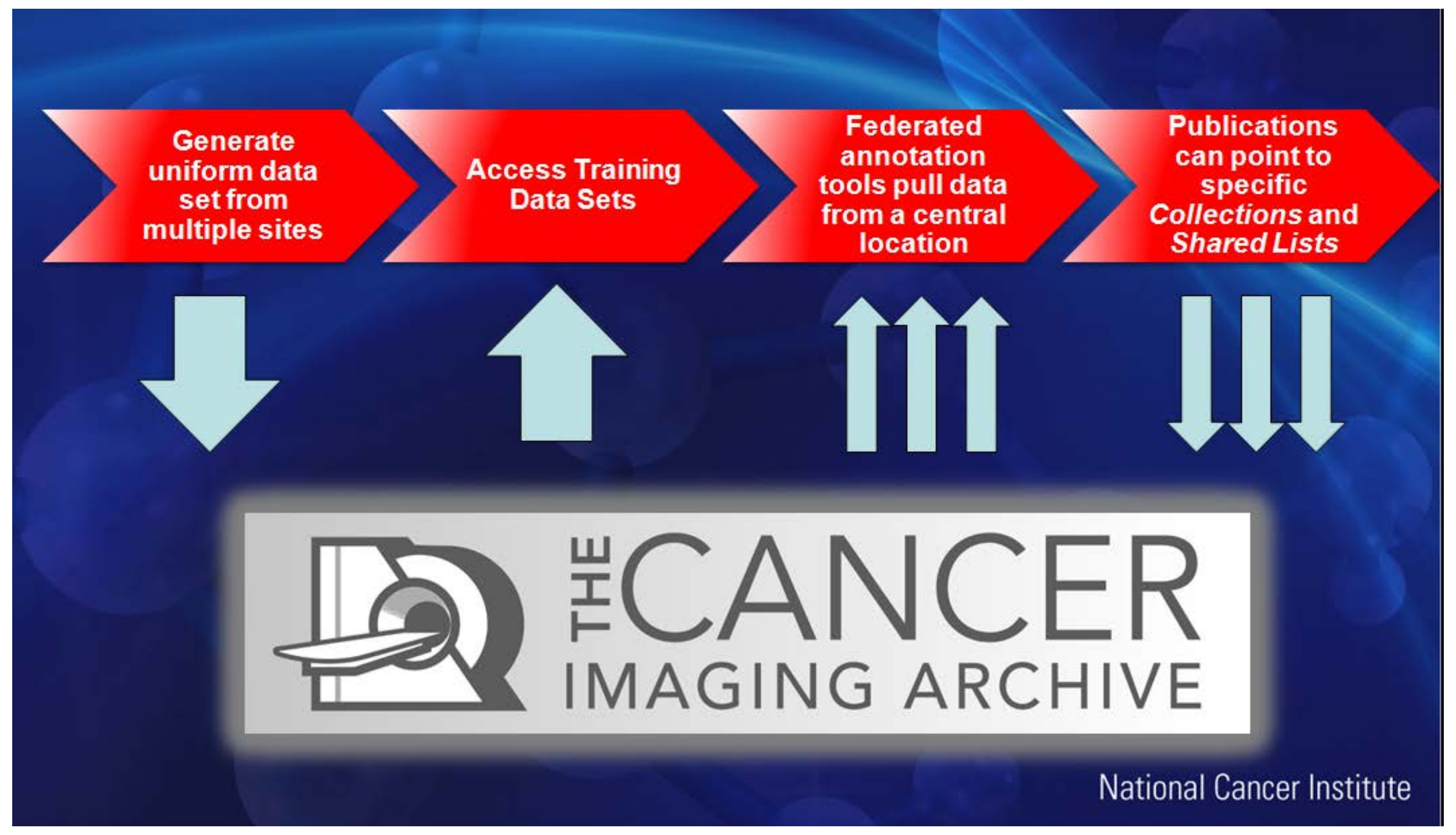

Generate uniform data set from multiple sites
Access Training Data Sets
Federated annotation tools pull data from a central location
Publications can point to specific *Collections* and *Shared Lists*
THE CANCER IMAGING ARCHIVE
National Cancer Institute

FIGURE 2: **TCIA provides high quality, fully de-identified image data that are linked through a common research participant ID to other TCGA data sets. Publications can point to specific TCIA collections.**

- **TCGA Renal Phenotype Research Group** —TCGA Renal Phenotype Research Group is part of the CIP TCGA Radiology Initiative. This activity is currently in the early stage of development. Multiple modalities of images which correlate to the kidney renal clear cell carcinoma (KIRC) data in the TCGA Data Portal are currently being gathered for submission to TCIA. In the mean time the group is beginning preliminary discussions around research methods and goals.

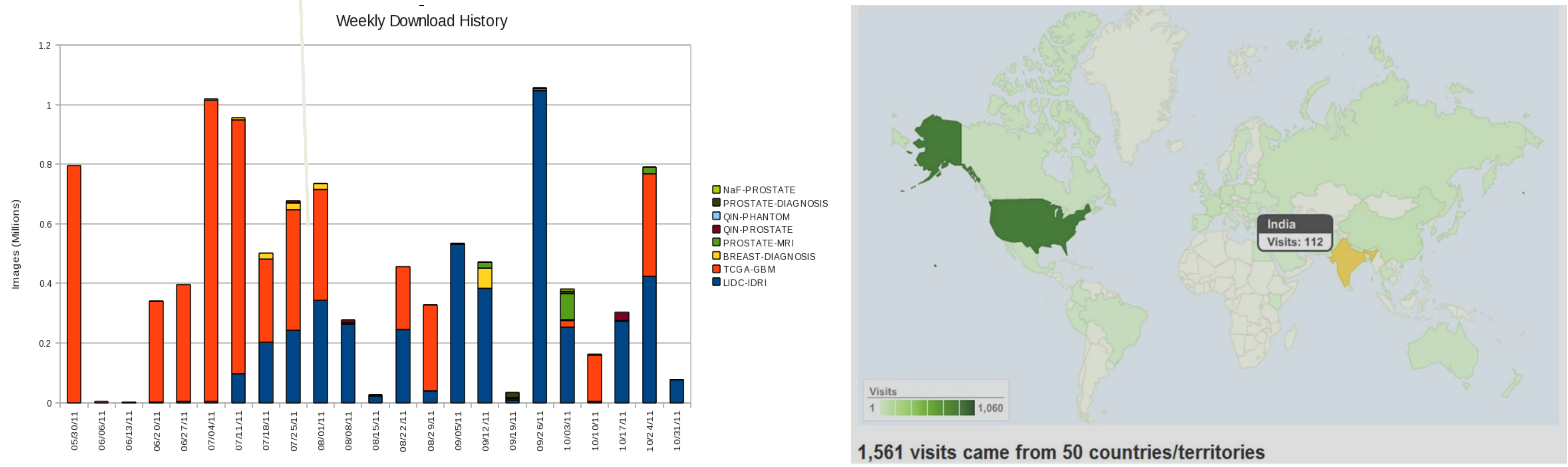


FIGURE 5. Approximately 10.5 million images have been downloaded by researchers, educators, and cancer patients in 50 countries around the world. TCIA has approximately 400 registered users and supports multiple active research initiatives.

CONCLUSIONS

TCIA is an actively managed information resource supported by a professional staff of image experts. The image repository supports all DICOM information objects. An integrated wiki hosts metadata and project descriptions. The system is hosted on a redundant, scalable hardware platform that ensures 99.9 % system availability. An IRB-Approved de-identification and curation workflow supported by a dedicated team ensures high quality data and full compliance with HIPAA and the Common Rule.

TCIA adds a new dimension to the Cancer Genome Atlas research program by enabling research on new quantitative and qualitative analyses that link imaging features to genomic signatures. TCIA also facilitates verification and validation of new computer aided analysis tools and imaging based candidate biomarkers.

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. HHSN261200800001E. 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.