

The Cancer Imaging Archive (TCIA)

De-Identification and HIPAA compliance procedure

The NCI Cancer Imaging Program has established the The Cancer Imaging Archive (TCIA) as its in-vivo imaging archive resource. The TCIA provides standards-based curation support to ensure safe and thorough de-identification of all images in the archive per federal HIPAA and HITECH regulations. In order to achieve this compliance without stripping the data of its scientific utility TCIA staff perform a redundant, thorough de-identification and analysis procedure based on guidance provided by the industry experts in DICOM standards committee Working Group 18.

DICOM standards committee Working Group 18 Supplement 142 provides a standard for image de-identification and a process with which to reduce the complexity involved in safely de-identifying DICOM image data while providing flexibility for scenarios which necessitate preservation of certain information needed for quality control and analysis that is essential to research. This is achieved by providing a number of Application Level Confidentiality Profiles which includes a Basic Profile along with a number of Option Profiles. These profiles provide the necessary instructions for how to safely clean DICOM elements which may contain PHI. The full Supplement 142 guidance document can be obtained at ftp://medical.nema.org/medical/dicom/final/sup142_ft.doc.

TCIA utilizes the RSNA Clinical Trials Processor (CTP) software in conjunction with caBIG's National Biomedical Imaging Archive (NBIA) to de-identify and host the images in the archive. The Cancer Imaging Program's Informatics Team has been working closely with the developer of CTP since 2009 to incorporate support for this standard as it was being defined by WG18. A full summary and time line of this project can be found at <https://wiki.nci.nih.gov/display/CIP/Incorporation+of+DICOM+WG18+Supplement+142+into+CTP>.

CTP provides an interface that allows application of any combination of the profiles to a set of images, and allows for application of an audit trail for retroactively tracking applied de-identification. For images that are submitted to TCIA the staff begins with the Basic Application Confidentiality Profile (which is the most aggressive) in combination with the following options:

- **Clean Descriptors Option:** Removal of identification information from descriptive tags which contain unstructured plain text values over which an operator has control
- **Retain Modified Longitudinal Temporal Information Options:** Modification of tags that contain dates or times
- **Retain Patient Characteristics Option:** Retention of physical characteristics of the patient that are descriptive rather than identifying information (e.g. metabolic measures, body weight, etc.)
- **Retain Device Identity Option:** Retention of information about the characteristics of the device used to perform the acquisition

- **Retain Safe Private Option:** Retention of Private Attributes confirmed not to contain PHI

After initial testing TCIA image curators individually inspect every image, both in the DICOM tags and the image pixels to ensure there is no PHI. Changes to the de-identification procedure are made as appropriate to correct any potential issues found by our curation team. After the completion of the image submissions the curation team again inspects every image in the full data set to ensure regulatory compliance. Only after this inspection is complete are the images made available to the general public.

For general information on what to expect as an image provider please see our web site at

<http://www.cancerimagingarchive.net/provider.html> or contact us directly at

CancerImagingArchive@mail.nih.gov.