De-identification Knowledge Base

Welcome!

This is the knowledge base referred to in

Moore SM, Maffitt DR, Smith KE, Kirby JS, Clark KW, Freymann JB, Vendt BA, Tarbox LR, Prior FW. De-identification of Medical Images with Retention of Scientific Research Value. RadioGraphics. 2015;35(3):727-35. doi: 10.1148/rg.2015140244

can be reached here: https://queries.cancerimagingarchive.net/PrivateElementKnowledgeBase/faces/index.xhtml

please be aware that this resource was last updated in 2015 ; Links to "mirgforge" are deprecated and no longer maintained.

Current practice per October 2019 can be found here: Submission and De-identification Overview

DICOM Basic Attribute Confidentiality Profile

The DICOM standards committee Working Group 18 (WG18) wrote Supplement 142, now incorporated into the published DICOM Standard. The Attribute Confidentiality Profile (DICOM PS 3.15: Appendix E) provides a standard for image de-identification, reducing the complexity involved in safely de-identifying DICOM image data while retaining the flexibility to preserve certain information for essential quality control and analysis. Application Level Confidentiality Profiles, which include a Basic Profile along with a number of Option Profiles, provide instructions for how to safely clean DICOM elements which may contain PHI. The DICOM Standard, including Part 15, is available at the NEMA web site. We recommend using the published standard above as it will be updated with any change proposals. We have also ported the contents of Table E.1-1 into XLS format for easy access.

Appendix E of PS 3.15 documents a system for protecting attributes; a section is reproduced below:

> *The Attributes listed in Table E.1-1 for each profile are contained in Standard IODs, or may be contained in Standard Extended IODs. An implementation claiming conformance to an Application Level*
>
> *Confidentiality Profile as a de-identifier shall protect or retain all instances of the Attributes listed in Table E.1-1, whether contained in the main dataset or embedded in an Item of a Sequence of Items. The following action codes are used in the table:*
>
> *– D – replace with a non-zero length value that may be a dummy value and consistent with the VR*
>
> *– Z – replace with a zero length value, or a non-zero length value that may be a dummy value and consistent with the VR*
>
> *– X – remove*
>
> *– K – keep (unchanged for non-sequence attributes, cleaned for sequences)*
>
> *– C – clean, that is replace with values of similar meaning known not to contain identifying information and consistent with the VR*
>
> *– U – replace with a non-zero length UID that is internally consistent within a set of Instances*
>
> *– Z/D – Z unless D is required to maintain IOD conformance (Type 2 versus Type 1)*
>
> *– X/Z – X unless Z is required to maintain IOD conformance (Type 3 versus Type 2)*
>
> *– X/D – X unless D is required to maintain IOD conformance (Type 3 versus Type 1)*
>
> *– X/Z/D – X unless Z or D is required to maintain IOD conformance (Type 3 versus Type 2 versus Type 1)*
>
> *– X/Z/U* - X unless Z or replacement of contained instance UIDs (U) is required to maintain IOD conformance (Type 3 versus Type 2 versus Type 1 sequences containing UID references)*

PS 3.15: E.2 then defines the Basic Application Level Confidentiality Profile, which describes how to apply the scheme above with a number of options that determine the level of protection provided. These definitions allow a system to follow a standard procedure and document the behavior of that system in a standard way.

DICOM Private Data Elements

We typically retain DICOM private data elements containing parameters that describe the acquisition while removing PHI. Performing this task requires understanding the mechanism defined by DICOM to support private elements. DICOM PS 3.5, section 7.8.1 states:

It is possible that multiple implementors may define Private Elements with the same (odd) group number. To avoid conflicts, Private Elements shall be assigned Private Data Element Tags according to the following rules.

a. *Private Creator Data Elements numbered (gggg,0010-00FF) (gggg is odd) shall be used to reserve a block of Elements with Group Number gggg for use by an individual implementor. The implementor shall insert an identification code in the first unused (unassigned) Element in this series to reserve a block of Private Elements. The VR of the private identification code shall be LO (Long String) and the VM shall be equal to 1.*

b. *Private Creator Data Element (gggg,0010), is a Type 1 Data Element that identifies the implementor reserving element (gggg,1000-10FF), Private Creator Data Element (gggg,0011) identifies the implementor reserving elements (gggg,1100-11FF), and so on, until Private Creator Data Element (gggg,00FF) identifies the implementor reserving elements (gggg,FF00-FFFF).*

c. *Encoders of Private Data Elements shall be able to dynamically assign private data to any available (unreserved) block(s) within the Private group, and specify this assignment through the blocks corresponding Private Creator Data Element(s). Decoders of Private Data shall be able to accept reserved blocks with a given Private Creator identification code at any position within the Private group specified by the blocks corresponding Private Creator Data Element.*

We use data in group 0009 as an example in the table below:

Tag	Description	Value
0009, 0010	Private Creator Element	ACME
0009, 1001	Average Density	15.5
0009, 1002	Density Standard Deviation	2.2

In the example, the element with tag (0009, 0010) is a private creator element with value "ACME". That reserves a block of elements for this manufacturer. The element (0009, 1001) is part of that block; the 10 in the element tag (1001) corresponds to the 10 in the tag of the Private Creator Element (0009, 0010).

This becomes complex when different manufacturers want to use the same reserved block to store information. When this occurs in a single image, the creator of the image reserves a block (for example, 0010). When a second application wants to add data to that same group, it detects the block written by the creator and creates a separate block (for example, 0011). The creator is not required to start at block 0010, but that appears to be common practice. The second or third application is not required to use 0011 or 0012. Based on this encoding scheme, some observations are:

1. If a collection of images is produced by equipment from different manufacturers, collisions may occur in the sets of private elements to be retained and discarded. For example, element (0009, 1001) from manufacturer A may contain an important physical parameter while that same element from manufacturer B may contain PHI.

2. If the collection has images created by an acquisition modality and then modified by another application (PACS, workstation), a private group may have multiple reserved blocks. Also, one cannot assume that the original creator will always choose reserved block 0010.

DICOM Tag Sniffer

- Download the software **NB link is deprecated
- Download the documentation **NB link is deprecated

To simplify implementation of "clean" instructions specified in DICOM PS 3.15, a new tool was developed to help inspect the contents of DICOM elements, allowing free text entry by a technician and Private Tags for potential PHI. This tool scans a folder, includes subfolders for DICOM objects, and produces several different outputs that depend on the mode used and input profiles. The software reads each DICOM object and iterates through each public and private element. The software then uses the profiles below to determine whether or not to retain the value of the element for later inspection:

- Confidentiality Profile: One input profile corresponds to the entries in table E.1-1 in DICOM PS 3.15. We list the attributes in the table and coded values according to table entries. When scanning DICOM objects, each public element is checked against the data in the profile. If the element is found in the profile, the software knows whether to record the element value for later inspection or ignore it. For example, if the DICOM profile indicates the element is to be deleted, there is no reason to review the value in that element.
- The Confidentiality Profile input is augmented with elements known to contain physical parameters such as rows, columns, or pixel spacing. Rather than tell the software to ignore values with a specific value representation, we list those elements explicitly.
- Modality Software Profile: This input profile describes private elements documented in the conformance statement by the manufacturer. This file takes into account the Private Creator Data Elements described above and has a code table for indicating program actions (record the value, ignore the value, etc).

These outputs are relevant at different stages of the curation and image publication process:

- Element Inventory: The element inventory consists of a set of DICOM tags found in the image set. The DICOM tags include only hexadecimal tags (xxxx, yyyy) and no values. All public and private tags are listed once. The Confidentiality Profile and Modality Software Profile are not consulted, as no values are retained for review.
- Element Values Before De-identification: Element values are examined to determine how to configure CTP scripts for proper de-identification. As mentioned above, we want to retain as many elements as possible while not exposing PHI. We also do not want to review all element values in all DICOM objects. We use a Confidentiality Profile that corresponds to the DICOM Basic Application Confidentiality Profile and a Modality Software Profile that properly describes private elements in DICOM objects.
- Element Values, Final Review: To review the values in DICOM objects after data is de-identified, as a final check before publication, this mode uses a different Confidentiality Profile and a different Modality Software Profile. For the Confidentiality Profile, we only list elements that are known physical parameters (rows, columns, etc.) and do not include the DICOM references from PS 3.15, Table E.1-1. That directs the software to record the element values. Likewise, the Modality Software Profile directs the software to record all values for later analysis.

This tool can be useful to the rest of the research community, so it has been made freely available as an open source application. We have also created documentation for its use in the context of a researcher’s own projects.

Private Element Knowledge Base Query Application

Data recorded in the documents above are also available through a web-based application with query capabilities. Researchers who obtain images through TCIA or by other means are welcome to search the database to find definitions for private elements.

- Private Element Knowledge Base Query Application

Manufacturer Specific Private Tags

As discussed above, medical manufacturers include private elements in their DICOM images to convey information not defined in the DICOM Standard. This section documents the information we have gathered from conformance statements.

The sections below describe information by manufacturer. That information is encoded in files that describe the private elements created by those manufacturers. Those files are part of the run-time environment of the Tag Sniffer and are maintained in our forge:

- https://mirgforge.wustl.edu/gf/project/dicomtagsniffer/scmsvn/?action=browse&path=%2Ftrunk%2Fdeploy%2Fprofiles%2Fdevice-profiles%2F

The information in the documents below is also available through a web based tool with query functions. That tool is found here:

- https://queries.cancerimagingarchive.net/PrivateElementKnowledgeBase/faces/index.xhtml

GE Medical Systems

GE Discovery CT	12/26/2012
GE Discovery MR	12/24/2012
GE Discovery PT	4/11/2012
GE HiSpeed CT	12/27/2012
GE HiSpeed LXiR0 CT	12/27/2012
GE HiSpeed QXi CT	12/27/2012
GE LightSpeed CT	7/12/2012
GE Signa MR series	3/28/2012

Philips

Philips Achieva MR series	12/30/2012
Philips Aura CT	7/17/2012
Philips Brilliance CT	7/16/2012

Siemens

Siemens CT	7/6/2012
Siemens Numaris MR	7/17/2012
Siemens Syngo MR	12/29/2012

Toshiba

Toshiba Aquilion CT	7/6/2012
Toshiba MR	7/18/2012

Software Tools

CTP

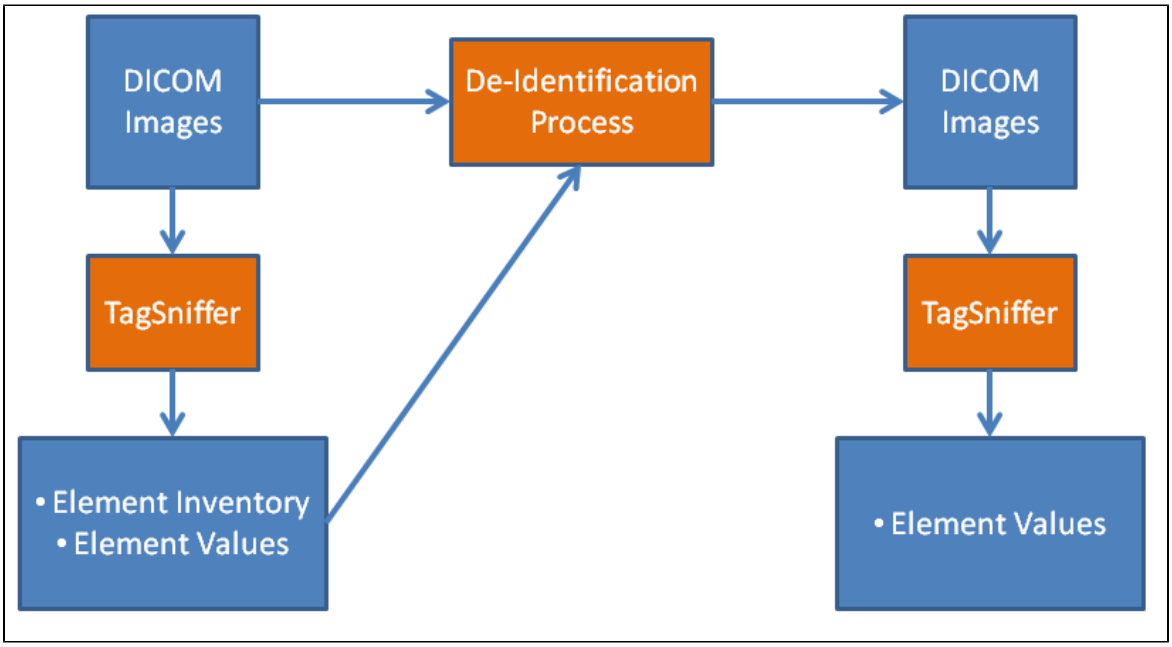
- Download the software
- Read the documentation
- Join the User Group mailing list
- Extending/Contributing CTP source code

TCIA utilizes the Radiological Society of North America (RSNA) 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. Find the summary and time line of this project here: 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 profiles to a set of images, and allows for an audit trail to retroactively track applied de-identification. When images 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

TCIA De-identification Work Flow



TCIA provides standards-based curation support to ensure safe and thorough de-identification of all images in the archive per federal HIPAA and Health Information Technology for Economic and Clinical Health (HITECH) Act regulations. To achieve this compliance without stripping the data of its scientific utility, TCIA performs a thorough de-identification and analysis procedure based on guidance from the DICOM standards committee WG18. Each collection submitted for publication is analyzed and de-identified as a whole using the steps listed below. All steps are completed before the collection is released for publication.

1. Each image in the collection is visually inspected to guarantee no PHI is burned into the pixel data.
2. Tag Sniffer is used to review the collection and produce an Element Inventory annotated with data from the DICOM Basic Application Confidentiality Profile and our set of Modality Software Profiles. This produces the list of DICOM elements found in the collection with a simple annotation scheme:
    a. One of the Basic Application Confidentiality Profile codes that indicates the DICOM scheme for de-identification (if the element is listed by DICOM).
    b. A simple code from our Modality Software Profile (No PHI: Retain, PHI: Delete, Not Sure: Review).
    c. No code, indicating the element is not registered.
3. Pre-identification output of the Tag Sniffer is also generated, containing the set of elements in the collection and all values that need to be reviewed for PHI. If the Basic Application Confidentiality Profile or applicable Modality Software Profile indicates the attribute is to be cleaned or that the attribute is a physical parameter that does not contain PHI, there is no need to review that element at this step. We know that our de-identification script will process the element properly.
4. Information from steps 2 and 3 is combined to create a CTP de-identification script for the collection. In the event of multiple scanners from different manufacturers, we might create and apply different scripts based on manufacturer.
5. The CTP de-identification script (or scripts) is (are) applied to the image collection and a separate copy of the images is created, retaining the original set in case we need to repeat a step.
6. Tag Sniffer is used to review the de-identified images and create the Final Review Output. This more complete output is reviewed by analysts to guarantee no PHI is carried forward after de-identification. Both public and private elements are included in the output for review.
7. If any errors are detected in de-identification in step 6, the CTP script is adjusted and the image set is processed again starting at step 5.

Only after this inspection is complete are the images made available to the general public. For information on what to expect as an image provider, please see our web site at https://www.cancerimagingarchive.net/primary-data/ and https://www.cancerimagingarchive.net/analysis-results/.

Background Information

Here are some presentations and papers which provide an overview on various aspects of DICOM de-identification and the official Supplement 142 de-identification standards:

1. Using RSNA's Clinical Trial Processor (CTP) Software for Clinical Trials and Research Applications, presentation, RSNA Annual Meeting, Chicago, IL, November 2012
2. Image Data Sharing for Biomedical Research: Meeting the De-identification and Informatics Challenges publication, Journal of Digital Imaging (DOI: 10.1007/s10278-011-9422-x)
3. Image Data Sharing for Biomedical Research: Meeting the De-identification and Informatics Challenges presentation, SIIM Annual Meeting, Washington, D.C., June 4, 2011
4. De-identification Revisited - DICOM Supplement 142 presentation, DICOM Conference 2010
5. Automated Standards-based Anonymization Profile for Image Sharing Using RSNA's Clinical Trial Processor poster with Q&A session, RSNA Annual Meeting, Chicago, IL, Nov 30, 2009