

PATIENT DE-IDENTIFICATION

Updated: 19 December 2018

Overview

Following industry best-practices, RECOMIA uses a standards-based approach to de-identification of DICOM images to insure that images are free of protected health information (PHI). Our de-identification process is developed in accordance with the requirements set forth by the Federal Drug Administration (FDA) and the European Medicines Agency (EMA). These requirements are defined the Health Insurance Portability and Accountability Act (HIPAA) section 164.514(b)(2) of the HIPPA Privacy Rule. The standard for de-identification of DICOM objects is defined by the DICOM Standard PS 3.15-2011 Digital Imaging and Communications in Medicine (DICOM), Part 15: Security and System Management Profiles.

The de-identification process in RECOMIA happens automatically and all data is de-identified locally at the submitting site before data it is uploaded via secure communication to the RECOMIA servers.

What is Protected Health Information (PHI)?

PHI is defined as "individually identifiable health information". In other words, information that can be used to directly or indirectly identify an individual in relation to the individual's past, present or future health condition and the provision of health care to the individual. Common types of PHI includes: patient name, address, birth date, social security number, medical and laboratory reports, physician name, hospital name, and date of examination. PHI can be embedded in both DICOM tags and pixel data.

RECOMIA de-identification process

The process of de-identification, by which PHI are removed from the health information in the data process by us, mitigates privacy risks to individuals and thereby supports the use of data for scientific research and other endeavours. The RECOMIA de-identification process is an automated 3-step process, in which two de-identification methods are deployed sequentially: 1) automated redaction of individual PHI identifiers (in DICOM tags and pixel data), and 2) formal determination by a qualified expert (i.e. only a qualified individual can determine when PHI has been properly removed). Both methods are run locally (in the client's web-browser) which means that no PHI will leave their closed network. Upon successful completion of the de-identification process, the de-identified data is automatically uploaded via secure communication to the RECOMIA servers.

Note, however, that both methods, even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds. Regardless, neither HIPAA, EMA, nor GDPR restricts the use or disclosure of de-identified health information, as it is no longer considered protected health information. Data processed by RECOMIA is an example of de-identified health information.

Step 1

Automated redaction of individual PHI identifiers stored in DICOM tags is the first step in our de-identification process. This step conform to the current DICOM standard to ensure data processed by us is transformed using approved reduction techniques such as generalisation of the data by grouping of values into categories, and suppression/masking of data where specific values, or whole records are removed from the dataset.

Step 2

The second step involves our optical character recognition (OCR) engine. In this step all images that are commonly known to store PHI (such as x-ray and mammography) is thoroughly scanned for characters embedded directly in the pixel data. This happens automatically, and in the event that PHI (or what our engine believes is PHI) is detected, the affected data will be invalidated. Invalidated data cannot be uploaded and requires manual expert review, as explained in step 3.

Step 3

Invalid data is automatically sent to manual expert review. During the review, the person responsible for data upload will get access to an OCR report detailing all detected characters. After careful consideration the invalid data can be either omitted from upload, manually validated by redaction (detected characters are blanked out), or accepted (in the event the discovered characters does not hold PHI) and in turn proceed to upload.

Step 1 to 3 is performed locally in the client' s web browser, and ensures no data containing PHI is uploaded to our servers. After successful de-identification, validated data is uploaded via secure communication to our servers.

Details

Base level de-identification

Patient Name and Patient ID are either blanked or modified. RECOMIA does not perform ID mapping between the original Patient ID and the ID that the images will have within RECOMIA. Any mapping that is performed manually at the submitting site, is the sole responsibility of the submitting site, and RECOMIA never sees the original Patient ID. Such data is defined as pseudo-de-identified data. To

show that the Patient Identity has been removed, the term “YES” is written into DICOM tag 00120062 “PatientIdentityRemoved” .

Exam Identifiers

DICOM makes extensive use of universal identifiers (UID) that could be used to identify a subject if a user had access to the PACS system at the institution where the images originated. RECOMIA uses its own root UID and then removes the original UID. UIDs have no special meaning other than serving as unique identifiers. This technique insures that images stay associated with the appropriate series, study, and subject as well as ensuring that referenced images between secondary capture images, structured reports, PET/CT, etc. are still valid references to images within RECOMIA.

Patient Demographics

The keep Patient Characteristics Option allows keeping some patient demographics for research purposes. The allowed fields are Patient’ s Sex, Patient’ s Age, Patient’ s Size, Patient’ s Weight, Ethnic Group, Smoking Status, and Pregnancy Status. If a subject is over 90 years of age, then the age must be listed as 90+. Allergies, Patient State (this is not where they live, rather their condition), Pre-Medication, and Special Needs are defined by the DICOM standard as “clean” and are kept by RECOMIA and examined for PHI along with all tags during curation. Other patient demographics such as birthdate, address, religious affiliations, etc. are removed or emptied.

Free Text

Free following free text fields are removed by RECOMIA during the curation process: Allergies, Patient State, Study Description, Series Description, Admitting Diagnoses Description, Admitting Diagnoses Code Sequence, Derivation Description, Identifying Comments, Medical Alerts, Occupation, Additional Patient History, Patient Comments, Contrast Bolus Agent, Protocol Name, Acquisition Device Processing Description, Acquisition Comments, Acquisition Protocol Description, Contribution Description, Image Comments, Frame Comments, Reason for Study, Requested Procedure Description, Requested Contrast Agent, Study Comments, Discharge Diagnosis Description, Service Episode Description, Visit Comments, Scheduled Procedure Step Description, Performed Procedure Step Description, Comments on Performed Procedure Step, Requested Procedure Comments, Reason for Imaging Service Request, Imaging Service Request Comments, Interpretation Text, Interpretation Diagnosis Description, Impressions, and Results Comments.

Devices

The Retain Device Identity Option of the DICOM de-identification standard allows for the retention of information related to the scanner used. The option allows for the following relevant tags to be retained: Station Name, Device Serial Number, Device UID, Plate ID, Generator ID, Cassette ID, Gantry ID, Detector ID, Scheduled Study Location, Scheduled Study Location AE Title, Scheduled Station AE Title, Scheduled Station Name, Scheduled Procedure Step Location, Performed Station AE Title, Performed Station Name, Performed Station Name Code Sequence, Scheduled Station Name Code

Sequence, Scheduled Station Geographic Location Code Sequence, and Performed Station Geographic Location Code Sequence. The tags listed above are retained if they are found to be free of PHI after RECOMIA curation of the submitted DICOM objects.

Private Tags

When a submitting site sends DICOM data to RECOMIA all private tags are removed.

Table 1

All odd group numbered tags are deleted. Table 1 details the de-identification performed for even grouped numbered tags at the submitting site by way of a RECOMIA supplied de-identification script.

TAG	Name	Value
0000, 1001	Requested SOPInstance UID	Replace with new randomly generated
0002, 0003	Media Storage SOPInstance UID	Replace with new randomly generated
0004, 1511	Referenced SOPInstance UIDIn File	Replace with new randomly generated
0008, 0014	Instance Creator UID	Replace with new randomly generated
0008, 0015	Instance Coercion Date Time	Delete
0008, 0018	SOPInstance UID	Add or replace with new randomly generated.
0008, 0050	Accession Number	Replace with new randomly generated
0008, 0058	Failed SOPInstance UIDList	Delete
0008, 0080	Institution Name	Delete
0008, 0081	Institution Address	Delete
0008, 0082	Institution Code Sequence	Delete
0008, 0090	Referring Physician Name	Replace with empty
0008, 0092	Referring Physician Address	Delete
0008, 0094	Referring Physician Telephone Numbers	Delete
0008, 0096	Referring Physician Identification Sequence	Delete
0008, 010D	Context Group Extension Creator UID	Replace with new randomly generated
0008, 1030	Study Description	Delete
0008, 103E	Series Description	Delete
0008, 1040	Institutional Department Name	Delete
0008, 1048	Physicians Of Record	Delete
0008, 1049	Physicians Of Record Identification Sequence	Delete

0008, 1050	Performing Physician Name	Delete
0008, 1052	Performing Physician Identification Sequence	Delete
0008, 1060	Name Of Physicians Reading Study	Delete
0008, 1062	Physicians Reading Study Identification Sequence	Delete
0008, 1070	Operators Name	Delete
0008, 1072	Operator Identification Sequence	Delete
0008, 1080	Admitting Diagnoses Description	Delete
0008, 1084	Admitting Diagnoses Code Sequence	Delete
0008, 1110	Referenced Study Sequence	Delete
0008, 1111	Referenced Performed Procedure Step Sequence	Delete
0008, 1120	Referenced Patient Sequence	Delete
0008, 1155	Referenced SOPInstance UID	Replace with new randomly generated
0008, 1195	Transaction UID	Replace with new randomly generated
0008, 2111	Derivation Description	Delete
0008, 2112	Source Image Sequence	Delete
0008, 3010	Irradiation Event UID	Replace with new randomly generated
0008, 4000	Identifying Comments	Delete
0008, 9123	Creator Version UID	Replace with new randomly generated
0010, 0010	Patient Name	Add or replace with new randomly generated.
0010, 0020	Patient ID	Add or replace with new randomly generated.
0010, 0021	Issuer Of Patient ID	Delete
0010, 0030	Patient Birth Date	Delete
0010, 0032	Patient Birth Time	Delete
0010, 0050	Patient Insurance Plan Code Sequence	Delete
0010, 0100	Referenced Patient Photo Sequence	Delete
0010, 0101	Patient Primary Language Code Sequence	Delete
0010, 0102	Patient Primary Language Modifier Code Sequence	Delete
0010, 1000	Other Patient IDs	Delete

0010, 1001	Other Patient Names	Delete
0010, 1002	Other Patient IDs Sequence	Delete
0010, 1005	Patient Birth Name	Delete
0010, 1040	Patient Address	Delete
0010, 1050	Insurance Plan Identification	Delete
0010, 1060	Patient Mother Birth Name	Delete
0010, 1080	Military Rank	Delete
0010, 1081	Branch Of Service	Delete
0010, 1090	Medical Record Locator	Delete
0010, 2000	Medical Alerts	Delete
0010, 2110	Allergies	Delete
0010, 2150	Country Of Residence	Delete
0010, 2152	Region Of Residence	Delete
0010, 2154	Patient Telephone Numbers	Delete
0010, 2180	Occupation	Delete
0010, 21B0	Additional Patient History	Delete
0010, 21F0	Patient Religious Preference	Delete
0010, 2297	Responsible Person	Delete
0010, 2299	Responsible Organization	Delete
0010, 4000	Patient Comments	Delete
0012, 0062	Patient Identity Removed	Add or replace with new randomly generated.
0012, 0063	De-identification Method	Add or replace with new randomly generated.
0018, 0010	Contrast Bolus Agent	Replace with empty
0018, 1030	Protocol Name	Delete
0018, 1400	Acquisition Device Processing Description	Delete
0018, 3100	Target UID	Replace with new randomly generated
0018, 4000	Acquisition Comments	Delete
0018, 9424	Acquisition Protocol Description	Delete
0018, A003	Contribution Description	Delete
0020, 000D	Study Instance UID	Replace with new randomly generated
0020, 000E	Series Instance UID	Replace with new randomly generated

0020, 0010	Study ID	Replace with empty
0020, 0052	Frame Of Reference UID	Replace with new randomly generated
0020, 0200	Synchronization Frame Of Reference UID	Replace with new randomly generated
0020, 3401	Modifying Device ID	Delete
0020, 3404	Modifying Device Manufacturer	Delete
0020, 3406	Modified Image Description	Delete
0020, 4000	Image Comments	Delete
0020, 9158	Frame Comments	Delete
0020, 9161	Concatenation UID	Replace with new randomly generated
0020, 9164	Dimension Organization UID	Replace with new randomly generated
0028, 1199	Palette Color Lookup Table UID	Replace with new randomly generated
0028, 1214	Large Palette Color Lookup Table UID	Replace with new randomly generated
0028, 4000	Image Presentation Comments	Delete
0032, 0012	Study ID Issuer	Delete
0032, 1030	Reason For Study	Delete
0032, 1032	Requesting Physician	Delete
0032, 1033	Requesting Service	Delete
0032, 1060	Requested Procedure Description	Delete
0032, 1070	Requested Contrast Agent	Delete
0032, 4000	Study Comments	Delete
0038, 0004	Referenced Patient Alias Sequence	Delete
0038, 0010	Admission ID	Delete
0038, 0011	Issuer Of Admission ID	Delete
0038, 001E	Scheduled Patient Institution Residence	Delete
0038, 0040	Discharge Diagnosis Description	Delete
0038, 0050	Special Needs	Delete
0038, 0060	Service Episode ID	Delete
0038, 0061	Issuer Of Service Episode ID	Delete
0038, 0062	Service Episode Description	Delete
0038, 0300	Current Patient Location	Delete
0038, 0400	Patient Institution Residence	Delete
0038, 0500	Patient State	Delete
0038, 4000	Visit Comments	Delete

0040, 0006	Scheduled Performing Physician Name	Delete
0040, 0007	Scheduled Procedure Step Description	Delete
0040, 000B	Scheduled Performing Physician Identification Sequence	Delete
0040, 0012	Pre Medication	Delete
0040, 0243	Performed Location	Delete
0040, 0253	Performed Procedure Step ID	Delete
0040, 0254	Performed Procedure Step Description	Delete
0040, 0275	Request Attributes Sequence	Delete
0040, 0280	Comments On The Performed Procedure Step	Delete
0040, 0555	Acquisition Context Sequence	Delete
0040, 1001	Requested Procedure ID	Delete
0040, 1004	Patient Transport Arrangements	Delete
0040, 1005	Requested Procedure Location	Delete
0040, 1010	Names Of Intended Recipients Of Results	Delete
0040, 1011	Intended Recipients Of Results Identification Sequence	Delete
0040, 1101	Person Identification Code Sequence	Delete
0040, 1102	Person Address	Delete
0040, 1103	Person Telephone Numbers	Delete
0040, 1400	Requested Procedure Comments	Delete
0040, 2001	Reason For The Imaging Service Request	Delete
0040, 2008	Order Entered By	Delete
0040, 2009	Order Enterer Location	Delete
0040, 2010	Order Callback Phone Number	Delete
0040, 2016	Placer Order Number Imaging Service Request	Replace with empty
0040, 2017	Filler Order Number Imaging Service Request	Replace with empty
0040, 2400	Imaging Service Request Comments	Delete
0040, 3001	Confidentiality Constraint On Patient Data Description	Delete

0040, 4023	Referenced General Purpose Scheduled Procedure Step Transaction UID	Replace with new randomly generated
0040, 4034	Scheduled Human Performers Sequence	Delete
0040, 4035	Actual Human Performers Sequence	Delete
0040, 4036	Human Performer Organization	Delete
0040, 4037	Human Performer Name	Delete
0040, A027	Verifying Organization	Delete
0040, A073	Verifying Observer Sequence	Delete
0040, A075	Verifying Observer Name	Replace with empty
0040, A078	Author Observer Sequence	Delete
0040, A07A	Participant Sequence	Delete
0040, A07C	Custodial Organization Sequence	Delete
0040, A088	Verifying Observer Identification Code Sequence	Delete
0040, A123	Person Name	Delete
0040, A124	UID	Replace with new randomly generated
0040, A171	Observation UIDTrial	Replace with new randomly generated
0040, A172	Referenced Observation UIDTrial	Replace with new randomly generated
0040, A307	Current Observer Trial	Delete
0040, A352	Verbal Source Trial	Delete
0040, A353	Address Trial	Delete
0040, A354	Telephone Number Trial	Delete
0040, A358	Verbal Source Identifier Code Sequence Trial	Delete
0040, A402	Observation Subject UIDTrial	Replace with new randomly generated
0040, A730	Content Sequence	Delete
0040, DB0C	Template Extension Organization UID	Replace with new randomly generated
0040, DB0D	Template Extension Creator UID	Replace with new randomly generated
0070, 0001	Graphic Annotation Sequence	Delete
0070, 0084	Content Creator Name	Replace with empty
0070, 0086	Content Creator Identification Code Sequence	Delete
0070, 031A	Fiducial UID	Replace with new randomly generated
0088, 0140	Storage Media File Set UID	Replace with new randomly generated

0088, 0200	Icon Image Sequence	Delete
0088, 0904	Topic Title	Delete
0088, 0906	Topic Subject	Delete
0088, 0910	Topic Author	Delete
0088, 0912	Topic Keywords	Delete
0400, 0100	Digital Signature UID	Delete
0400, 0402	Referenced Digital Signature Sequence	Delete
0400, 0403	Referenced SOPInstance MACSequence	Delete
0400, 0404	MAC	Delete
0400, 0550	Modified Attributes Sequence	Delete
0400, 0561	Original Attributes Sequence	Delete
2030, 0020	Text String	Delete
3006, 0024	Referenced Frame Of Reference UID	Replace with new randomly generated
3006, 00C2	Related Frame Of Reference UID	Replace with new randomly generated
300A, 0013	Dose Reference UID	Replace with new randomly generated
300E, 0008	Reviewer Name	Delete
4000, 0010	Arbitrary	Delete
4000, 4000	Text Comments	Delete
4008, 0042	Results IDIssuer	Delete
4008, 0102	Interpretation Recorder	Delete
4008, 010A	Interpretation Transcriber	Delete
4008, 010B	Interpretation Text	Delete
4008, 010C	Interpretation Author	Delete
4008, 0111	Interpretation Approver Sequence	Delete
4008, 0114	Physician Approving Interpretation	Delete
4008, 0115	Interpretation Diagnosis Description	Delete
4008, 0118	Results Distribution List Sequence	Delete
4008, 0119	Distribution Name	Delete
4008, 011A	Distribution Address	Delete
4008, 0202	Interpretation IDIssuer	Delete
4008, 0300	Impressions	Delete
4008, 4000	Results Comments	Delete
5000, 3000	Curve Data	Delete
6000, 3000	Overlay Data	Delete

6000, 4000	Overlay Comments	Delete
FFFA, FFFA	Digital Signatures Sequence	Delete
FFFC, FFFC	Data Set Trailing Padding	Delete