Appendix A: Data Considered for Re-identification Risk Measurement

The data considered for re-identification risk measurement is described in the following subsections.

A.1: Pixel Data

Pixel data must be screened for residual and marginal identifiers, with recordkeeping to support the findings of the screening process. Pixel data screening guidance is provided as part of the full report.

A.2: File Names and File Paths

File Names and File Paths should be transformed as needed to ensure no Direct Identifiers or Demographic Quasi-Identifiers are retained.

A.3: DICOM Headers

In the following sections, we classify several commonly used tags from the DICOM standard and provide de-identification recommendations. Some of these are specific to the Cohort pertaining to the image request.

Any tag not explicitly permitted in these subsections must be masked as part of the de-identification strategy.

A.3.1: Direct Identifiers

The following tags are classified as Direct Identifiers. They need to be *masked* during de-identification, by secure hashing, secure encryption, or being redacted.

DICOM TAG	TAG DESCRIPTION
(0008,0050)	Accession Number
(0010,0010)	Patient's Name
(0010,0020)	Patient ID
(0010,1000)	Other Patient IDs
(0010,1001)	Other Patient Names
(0010,1005)	Patient's Birth Name
(0010,1040)	Patient's Address
(0010,2154)	Patient's Telephone Numbers
(0010,2295)	Breed Registration Number
(0012,0020)	Clinical Trial Protocol ID
(0012,0030)	Clinical Trial Site ID

DICOM TAG	TAG DESCRIPTION
(0012,0040)	Clinical Trial Subject ID
(0012,0042)	Clinical Trial Subject Reading ID
(0012,0071)	Clinical Trial Series ID
(0018,9445)	*(no description - NEMA placeholder)*
(0020,0010)	Study ID
(0020,1000)	Series in Study
(0020,9056)	Stack ID
(0032,000A)	Study Status ID
(0032,000C)	Study Priority ID
(0032,0012)	Study ID Issuer
(0038,0008)	Visit Status ID

DICOM TAG	TAG DESCRIPTION
(0038,0010)	Admission ID
(0038,0400)	Patient's Institution Residence
(0040,0031)	Local Namespace Entity ID
(0040,0032)	Universal Entity ID
(0040,0033)	Universal Entity ID Type
(0040,2016)	Placer Order Number / Imaging Service Request
(0040,2017)	Filler Order Number / Imaging Service Request
(0040,A123)	Person Name
(0070,0080)	Content Label
(0400,0005)	MAC ID Number
(0400,0020)	Data Elements Signed
(0400,0564)	Source of Previous Values
(300A,0182)	Patient Setup Number
(4008,0040)	Results ID
(4008,0119)	Distribution Name
(4008,011A)	Distribution Address
(4008,0200)	Interpretation ID
(4008,0210)	Interpretation Type ID
(4008,0212)	Interpretation Status ID

The following tags are Unique Image Identifiers, which tie directly to an image and thus a patient record. As such, they are also classified as Direct Identifiers of the patients. They also need to be *masked* during de-identification, by secure hashing, secure encryption, or being redacted.

DICOM TAG	TAG DESCRIPTION
(0008,0014)	Instance Creator UID
(0008,0018)	SOP Instance UID
(0008,010C)	Coding Scheme UID
(0008,010D)	Context Group Extension Creator UID
(0008,1150)	Referenced SOP Class UID
(0008,1155)	Referenced SOP Instance UID
(0008,3010)	Irradiation Event UID
(0008,9123)	Creator-Version UID
(0020,000D)	Study Instance UID
(0020,000E)	Series Instance UID
(0020,0052)	Frame of Reference UID
(0020,0200)	Synchronization Frame of Reference UID
(0020,9164)	Dimension Organization UID
(0040,A124)	UID
(0088,0140)	Storage Media File-set UID
(0400,0010)	MAC Calculation Transfer Syntax UID
(0400,0100)	Digital Signature UID
(3006,0024)	Referenced Frame of Reference UID
(3006,00C2)	Related Frame of Reference UID

A.3.2: Demographic Quasi-Identifiers

This section lists the Demographic Quasi-Identifiers in the DICOM tags. They need to be transformed according to the recommendations of our statistical analysis. Some fields such as Marital Status and Patient Location have consistent transformations; others such as Date of Birth, Ethnicity / Race, and Gender are transformed according to the recommendations of the Cohort.

Transformations (or reference to the Cohort definition) are provided below.

QI GROUP	RECOMMENDED TRANSFORMATIONS	DICOM TAG	TAG DESCRIPTION
Date of Birth	Apply Cohort-specific age generalization interval	(0010,0030)	Patient's Birth Date
		(0010,1010)	Patient's Age
Ethnicity / Race	All Cohorts: - Redact ethnicity information (i.e., Hispanic/Latino status), and - Apply Cohort-specific race grouping	(0010,2160)	Ethnic Group
Gender	Pediatric Cohort: - Removal of "Unknown" value All other Cohorts: - Retain	(0010,0040)	Patient's Sex
		(0010,2203)	Patient's Sex Neutered
Marital Status	All Cohorts: - Redact	(0010,2298)	Responsible Person Role

Patient Location	All Cohorts:
	- Locations outside the US:
	  o Redact all location information pertaining to patients
	- For US Locations:
	  o Truncation of ZIP Code to 3 digits, and removal of ZIP codes with less than 20,000 inhabitants (036, 692, 878, 059, 790, 879, 063, 821, 884, 102, 823, 890, 203, 830, 893, 556, and 831)
	  o Redaction of State and City
		(0008,0201)	Timezone Offset From UTC

		(0010,2150)	Country of Residence
		(0010,2152)	Region of Residence
		(0038,0300)	Current Patient Location

Provider Location	All Cohorts:
	- Locations outside the US:
	  o Redact all location information pertaining to patients
	- For US Locations:
	  o Truncation of ZIP Code to 3 digits, and removal of ZIP codes with less than 20,000 inhabitants (036, 692, 878, 059, 790, 879, 063, 821, 884, 102, 823, 890, 203, 830, 893, 556, and 831)
	Redaction of State and City
		(0008,0054)	Retrieve AE Title

		(0008,0080)	Institution Name
		(0008,0081)	Institution Address
		(0008,0090)	Referring Physician's Name
		(0008,0092)	Referring Physician's Address
		(0008,0094)	Referring Physician's Telephone Numbers
		(0008,1010)	Station Name
		(0008,1048)	Physician(s) of Record
		(0008,1050)	Performing Physician's Name
		(0008,1060)	Name of Physician(s) Reading Study
		(0008,1070)	Operators' Name
		(0010,0021)	Issuer of Patient ID
		(0010,2299)	Responsible Organization
		(0012,0010)	Clinical Trial Sponsor Name
		(0012,0031)	Clinical Trial Site Name

QI GROUP	RECOMMENDED TRANSFORMATIONS	DICOM TAG	TAG DESCRIPTION
		(0012,0060)	Clinical Trial Coordinating Center Name
		(0018,1000)	Device Serial Number
		(0018,1250)	Receive Coil Name
		(0018,1251)	Transmit Coil Name
		(0038,0011)	Issuer of Admission ID
		(0040,0001)	Scheduled Station AE Title
		(0040,0010)	Scheduled Station Name
		(0040,0035)	Identifier Type Code
		(0040,0241)	Performed Station AE Title
		(0040,0242)	Performed Station Name
		(0040,1010)	Names of Intended Recipients of Results
		(0040,2008)	Order Entered By
		(0040,2009)	Order Enterer's Location
		(0040,2010)	Order Callback Phone Number
		(0040,A075)	Verifying Observer Name
		(0070,0084)	Content Creator's Name
		(0088,0130)	Storage Media File-set ID
		(0400,0115)	Certificate of Signer
		(0400,0120)	Signature
		(3006,00A6)	ROI Interpreter
		(4008,010A)	Interpretation Transcriber
		(4008,010C)	Interpretation Author
		(4008,0114)	Physician Approving Interpretation

The Cohort-specific transformations are described below:

VARIABLE	COHORT-SPECIFIC TRANSFORMATIONS			
	PEDIATRIC	HEAD, CANCER/AUTO-IMMUNE	ORTHOPEDIC	THORACIC
Date of Birth	Year of Birth in 3-year intervals from 1934	Year of Birth in 2-year intervals from 1934, 1933 or older in a group	Year of Birth in 3-year intervals from 1934, 1933 or older in a group	Year of Birth in 5-year intervals from 1934, 1933 or older in a group

Gender	Removal of "Unknown" value	Retain
Race	Consolidate into: White, Black, Asian, Other	Consolidate into: Pacific Islander, American Indian or Native Indian, Asian, Black or African American, Other, White

A.3.3: Free-text and Non-standardized Fields

Free-text and non-standardized fields need to be redacted due to the inconsistent and uncharacterized nature of their contents in the general case. The free text and non-standardized Private Band Tags are listed below.

DICOM TAG	TAG DESCRIPTION
(0008,0100)	Code Value
(0008,0102)	Coding Scheme Designator
(0008,0103)	Coding Scheme Version
(0008,0104)	Code Meaning
(0008,0105)	Mapping Resource
(0008,0106)	Context Group Version
(0008,0107)	Context Group Local Version
(0008,010B)	Context Group Extension Flag
(0008,010F)	Context Identifier
(0008,0115)	Coding Scheme Name
(0008,0116)	Coding Scheme Responsible Organization
(0008,0117)	Context UID
(0008,1030)	Study Description
(0008,103E)	Series Description
(0008,1080)	Admitting Diagnoses Description
(0008,2111)	Derivation Description
(0008,4000)	Identifying Comments
(0010,1060)	Patient's Mother's Birth Name
(0010,1080)	Military Rank
(0010,1081)	Branch of Service
(0010,2180)	Occupation
(0010,1090)	Medical Record Locator
(0010,2000)	Medical Alerts
(0010,2110)	Allergies
(0010,21B0)	Additional Patient History

DICOM TAG	TAG DESCRIPTION
(0010,21F0)	Patient's Religious Preference
(0010,2201)	Patient Species Description
(0010,2292)	Patient Breed Description
(0010,4000)	Patient Comments
(0012,0021)	Clinical Trial Protocol Name
(0012,0050)	Clinical Trial Time Point ID
(0012,0051)	Clinical Trial Time Point Description
(0012,0072)	Clinical Trial Series Description
(0018,1030)	Protocol Name
(0018,1400)	Acquisition Device Processing Description
(0018,4000)	Acquisition Comments
(0018,7006)	Detector Description
(0018,7062)	Exposure Control Mode Description
(0018,9080)	Metabolite Map Description
(0018,A003)	Contribution Description
(0020,4000)	Image Comments
(0020,9421)	Dimension Description Label
(0028,0A04)	Pixel Spacing Calibration Description
(0028,3003)	LUT Explanation
(0028,6022)	Frame of Interest Description
(0032,1030)	Reason for Study
(0032,1060)	Requested Procedure Description
(0032,4000)	Study Comments

DICOM TAG	TAG DESCRIPTION	DICOM TAG	TAG DESCRIPTION
(0038,0040)	Discharge Diagnosis Description	(0042,0010)	Document Title
(0038,0050)	Special Needs	(0042,0011)	Encapsulated Document
(0038,0500)	Patient State	(0054,0018)	Energy Window Name
(0038,4000)	Visit Comments	(0054,0039)	Phase Description
(0040,0007)	Scheduled Procedure Step Description	(0054,0400)	Image ID
(0040,0009)	Scheduled Procedure Step ID	(0070,0006)	Unformatted Text Value
(0040,0012)	Pre-Medication	(0070,0068)	Graphic Layer Description
(0040,0253)	Performed Procedure Step ID	(0070,0081)	Content Description
(0040,0254)	Performed Procedure Step Description	(0088,0912)	Topic Keywords
(0040,0255)	Performed Procedure Type Description	(3006,0002)	Structure Set Label
		(3006,0006)	Structure Set Description
(0040,0280)	Comments on the Performed Procedure Step	(3006,0026)	ROI Name
(0040,0310)	Comments on Radiation Dose	(3006,0028)	ROI Description
(0040,0400)	Comments on the Scheduled Procedure Step	(4000,4000)	Text Comments
		(4008,010B)	Interpretation Text
(0040,1001)	Requested Procedure ID	(4008,0115)	Interpretation Diagnosis Description
(0040,1002)	Reason for the Requested Procedure	(5000,0022)	Curve Description
(0040,1400)	Requested Procedure Comments	(5002,0022)	Curve Description
(0040,2001)	Reason for the Imaging Service Request	(6000,0022)	Overlay Description
		(6000,1500)	Overlay Label
(0040,2400)	Imaging Service Request Comments	(6002,0022)	Overlay Description
		(6002,1500)	Overlay Label
(0040,3001)	Confidentiality Constraint on Patient Data Description	Any tag with an odd 4th digit (i.e., 1, 3, 5, 7, 9, B, C, F)	DICOM Private Tag
(0040,A160)	Text Value		
(0040,A492)	Completion Flag Description		

A.3.4: Non-identifiers

Non-identifiers do not contribute meaningful identifiability under the release scenarios considered in the report. They can be maintained in the de-identified data without any modification to the field contents.