Lung Image Database Consortium's Normalized Relational Database

- De-identified (anonymized) CT datasets
- Publicly accessible
- Web search tools
- Cases from search results can be downloaded
  - ftp or DICOM

De-identified CT datasets

Digital Imaging and Communications in Medicine (DICOM)

Supplement 55: Attribute Level Confidentiality (including De-identification)

DICOM Standards Committee, Working Group 14, Security, VERSION: Public Comment, 15 May 2001

Scrubbed Attribute Name Tag (group,element)

Instance Creator UID (0008,0014)
SOP Instance UID (0008,0018)
Accession Number (0008,0050)
Institution Name (0008,0080)
Institution Address (0008,0081)
Referring Physician's Name (0008,0090)
Station Name (0008,1010)
Study Description (0008,1030)
Series Description (0008,103E)
Institutional Department Name (0008,1040)
Physician(s) of Record (0008,1048)
Performing Physicians' Name (0008,1050)
Name of Physician(s) Reading Study (0008,1060)
Operators' Name (0008,1070)
Admitting Diagnoses Description (0008,1080)
Referenced SOP Instance UID (0008,1155)
Derivation Description (0008,2111)
Patient's Name (0010,0010)
Patient ID (0010,0020)
Patient's Birth Date (0010,0030)
Patient's Birth Time (0010,0032)
Patient's Sex (0010,0040)
Other Patient Ids (0010,1000)
Other Patient Names (0010,1001)
Patient's Age (0010,1010)
Patient's Size (0010,1020)
Patient's Weight (0010,1030)
Ethnic Group (0010,2160)
Occupation (0010,2180)
Additional Patient's History (0010,21B0)
Patient Comments (0010,4000)
Device Serial Number (0018,1000)
Protocol Name (0018,1030)
Study Instance UID (0020,000D)
Series Instance UID (0020,000E)
Study ID (0020,0010)
Frame of Reference UID (0020,0052)
Synchronization Frame of Reference UID (0020,0200)
Image Comments (0020,4000)
UID (0040,A124)

Scrubbing

- "On the fly"
  - No intermediate records written to disk files
    - » Modified Mallinckrodt DICOM Central Test Node (CTN) Software, Version 3.0.3
    - » Implemented using large memory buffers (2GB) such that existing CTN utility subroutines could be used, e.g. DCM_Modify_Elements, before buffer contents are written to disk
    - » Assistance from Stephen Moore, primary author

Research Case ID ↔ Patient ID
(plain text) (encrypted)

- Sequentially generated "Research Case ID" replaces Patient ID in modified DICOM header
- One-to-one relationship between Research Case ID and Patient ID preserved in encrypted table on backend machine.
- Table decrypted as necessary to support continued data entry
  - New CT exams
  - Pathology coded by Patient ID
  - Database QC

Automatic database loading of modified DICOM header info

- Supports automatic construction of
  - Research Case ID Table
  - Study Table
  - Series Table
- Immediacy
  - Database is publicly available the instant the DICOM format data is sorted
  - Records updated as modified by Experts/Pathology.
- MySQL, PHP, Apachie

Research Case ID table

research case						
case Id (autoincr)	encrypted institutionRecordId	rel Institution Id	rel Ethnicity Id (study entry form)	Gender (0010 0040)	relSmoking History (from medical record history)	pack years
1	agj(&@#...	5	1	F	1	0
2	is^#&#...	5	4	M	2	52
3	^&*^^^Kjj6...	5	5	M	4	300

CT Study table

study								
studyId (auto incr)	encrypted Study Instance UID (0020 000d)	rel Case Id	Study Date: YYYYMMDD (0008: 0020)	rel Modality Id (0008 0060)	relManu-facturer Id (0008 0070)	model Name (0008 1090)	study Descriptive Text	subject Age (0010 1010)
1	%$^&(*&(9	1	20020717	1	1	LSUltra	Lung	55
2	$#!$#$*)0*(*&	1	20030123	1	1	LSUltra	Lung	31
3	%$&#@&88*()^	2	20030405	4	4	Emotion Duo	PET-CT Thorax	43

CT Series table

seriesId (auto incr)	encrypted Series Instance UID (0020 000e)	relStudyId	relStorageDirectory Id (identifies loci of stored series' image files)	relSeries Description Id (radiologist definition session entry)	lung field: 1: full 2: partial (radiologist definition session entry)
1	)(*&%$%^%$	1	1	1	1
2	&*^*%$	1	1	3	1
3	)(**^$#FHHG	1	1	10	1
…	…	…	…	…	…
990	*(&HGFDT$R	2	2	2	1
...	...	...	...	...	...
4558	)*GDXDS$$^	3	4		

CT Series table, continued

acq&ReconParams																				
Contrast / Bolus Agent (0018 0010)	Scan Options, e.g. Helical, etc. (0018 0022)	slice thickness (0018 0050)		KVP (0018 0060)	Spacing between slices (0018 0088)		Data Collecion Diameter (0018 0090)	Software Version (0018 1020)	Contrast/ Bolus Route (0018 1040)	Reconstruction Diameter (0018 1100)	Source-Detector Distance (0018 1110)	Source-subject Distance (0018 1111)	Gantry Tilt (0018 1120)	Exposure Time (0018 1150)	Tube Current (0018 1151)	Exposure (0018 1152)	Extremety Filter Type (0018 1160)	Focal Spot Size (0018 1190)	Reconstruction Kernel (0018 1210)	Subject Position (0018 5100)
		min	max		min	max														
…	…		…	…		…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
...	...		...	...		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Series Meeting Inclusion Criteria

objects identified in a series			
seriesObjectId (auto incr)	relSeriesId	approxObjectLoc (x,y,z pixel coords)	approxTextualLociDescription (Text String: L/R Lung,Inf,Sup,Med,Lat,Ant,Post)
…	…	…	…
2045	990	300,100,400	RL,L,A,S (R Lung, Lat,Ant,Sup)
2046	990	300,300,100	RL,L,A,S
2047	991	150,255,50	LL,I

Nodule Expert Spatial Definitions

nodule spatial definitions												
nodule Spatial DefinitionId (auto incr)	relNoduleId	relSeries Object Id	relExpert Id	definedQuantitativeParams…			relObjectTy pe	relNodule Class	noduleMargins			"action-able" (1-5)
				centroid locus (I,j,k)	volume, cu mm	definitionFileId (format not defined yet)			poorly defnied (1-5)	spiculated (1-5)	smooth (1-5)	
…	…		…	…	…	…	…					
121	1	2045	1	35,-47,256	9	…/D0013346	N	1	3	2	4	
122	1	2045	2	36,-47,254	11	…/D0013347	N	2	3	2	4	
123	1	2045	3	35,-45,255	12	…/D0013348	SL					
124	1	2045	4	35.3,-45.6,254	10.6	…/P0013						
…	…	…	…	…	…	…						
578	1	10476	2	37,-48,255	9.3	…/D0203335						

Pathology/Cytology Results

nodulePathology			
nodulePathId	relNoduleId	Date YYYYMMDD	relPathology DiagnosisId
1	1	20011029	3
2	1	20020516	5
…	…	…	…
45	39	20020711	12
…	…	…	…
147	1	20020625	6

(Very) Small Sample of Pathology table NCICB's Common Data Elements

T0	No evidence of primary tumor
Tis	Carcinoma in situ
T1	Tumor 3 cm or less in greatest dimension, surrounded by lung or visceral pleura, invasion more proximal than the lobar bronchus, (i.e., not in the main bronchus)
T2	Tumor with any of the following features of size or extent: More than 3 cm in greatest dimension Invades the visceral pleura Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not invole the entire lung
T3	Tumor of any size that directly invades any of the following: chest wall (including superior sulcus tumors), diaphragm, mediastinal pleura, parietal pericardium; or tumor in the main bronchus less than 2 cm distal to the carina, but without involvement of
T4	Tumor of any size that invades any of the following: mediastinum, heart, great vessels, trachea, esophagus, vertebral body, carina; or separate tumor nodules in the same lobe; or tumor with a malignant pleural effusion

For the LIDC

NCICB

- Denise Warzel

Univ. Chicago

- Sam Armato
- Roger Engelmann

UCLA

- Mike McNitt-Gray

Univ. Michigan

- Gary Laderach
- Peyton Bland
- Chuck Meyer