# Standardized representation of the LIDC annotations using DICOM

● ● ●

Andrey Fedorov, Matthew Hancock, David Clunie,
Mathias Brockhausen, Jonathan Bona, Justin Kirby, John Freymann,
Steve Pieper, Hugo Aerts, Ron Kikinis, Fred Prior
7 January 2019 for CBIIT Imaging Community
Contact: andrey.fedorov@gmail.com

# Lung Imaging Database Consortium (LIDC)

- NCI-initiated in 2000, 5 institutions selected through peer-review process
- Goal: develop a public web-accessible resource containing
  - an image repository of screening and diagnostic thoracic CT scans
  - associated metadata such as technical scan parameters (e.g., slice thickness, tube current, and re- construction algorithm and patient information (e.g., age, gender, and pathologic diagnosis)
  - nodule truth information based on the subjective assessments of mul- tiple experienced radiologists (e.g., lesion category, nodule outlines, and subtlety ratings)

> Although the development of CAD methods for lung nodules on CT scans has accelerated, all of these diverse approaches have one common constraint: Access to well-characterized image data is limited. This deficiency represents a fundamental limitation to CAD research in this area. Consequently, it is difficult to develop and test CAD methods in a robust and reliable fashion.

Armato et al. Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community1. *Radiology* **232,** 739–748 (2004).
Armato et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **38,** 915–931 (2011).

# TCIA LIDC-IDRI collection

- One of the largest and most popular annotated collection of TCIA
- Utilized in numerous challenges
- 1000+ citations, used to enable numerous derivative studies

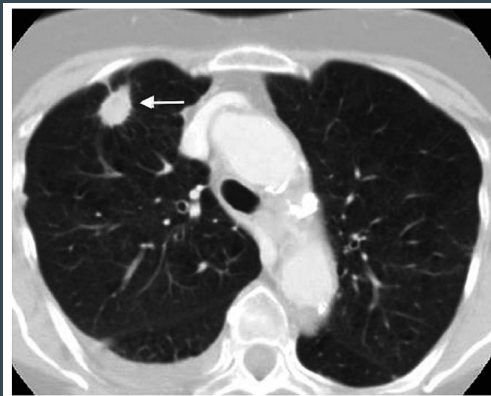https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

Kalpathy-Cramer, J., Freymann, J. B., Kirby, J. S., Kinahan, P. E. & Prior, F. W. Quantitative Imaging Network: Data Sharing and Competitive AlgorithmValidation Leveraging The Cancer Imaging Archive. *Transl. Oncol.* **7**, 147–152 (2014).

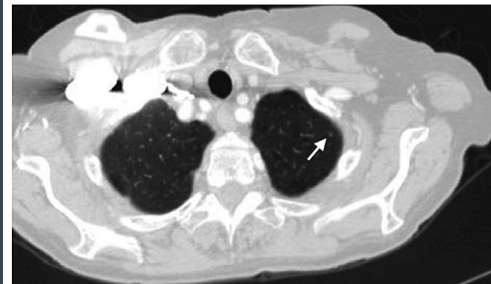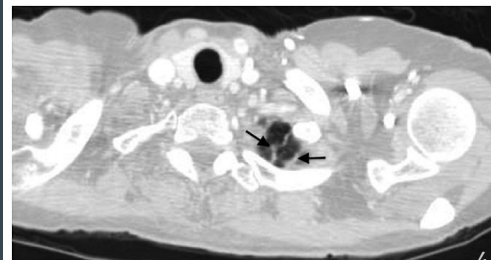| Collection | Cancer Type | Location | Subjects | Image Types | Supporting Data | Access |
|---|---|---|---|---|---|---|
| National Lung Screening Trial | Lung Cancer | Chest | 26254 | CT, Pathology | Clinical | Limited |
| CBIS-DDSM | Breast Cancer | Breast | 6671 | MG | Image Analyses | Public |
| LIDC-IDRI | Lung Cancer | Chest | 1010 | CT, CR, DX | Image Analyses, Clinical | Public |
| CT Colonography | Colon Cancer | Colon | 825 | CT | Image Analyses | Public |



3
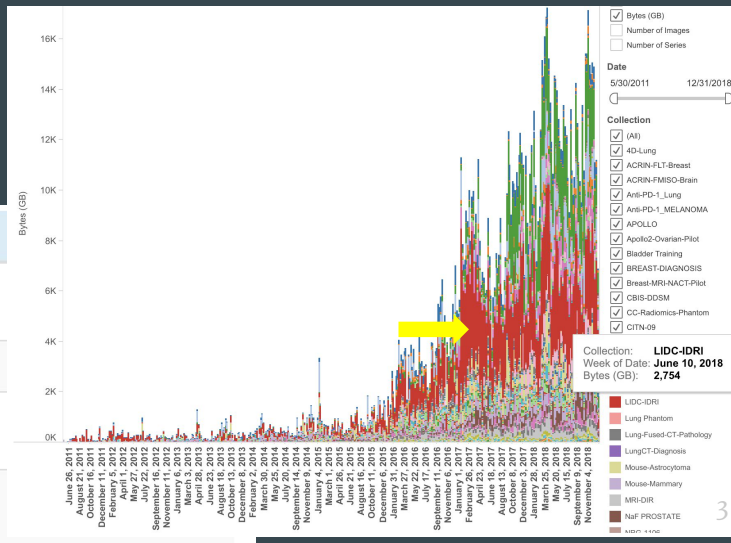
# LIDC groups of findings



1. **"nodules ≥ 3 mm"**: any lesion considered to be a nodule with greatest in-plane dimension in the range 3–30 mm regardless of presumed histology;

2. **"nodules < 3 mm"**: any lesion considered to be a nodule with greatest in-plane dimension less than 3 mm that is not clearly benign;

3. **"non-nodules ≥ 3 mm"**: any other pulmonary lesion, such as an apical scar, with greatest in-plane dimension greater than or equal to 3 mm that does not possess features consistent with those of a nodule.

Armato et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **38,** 915–931 (2011).
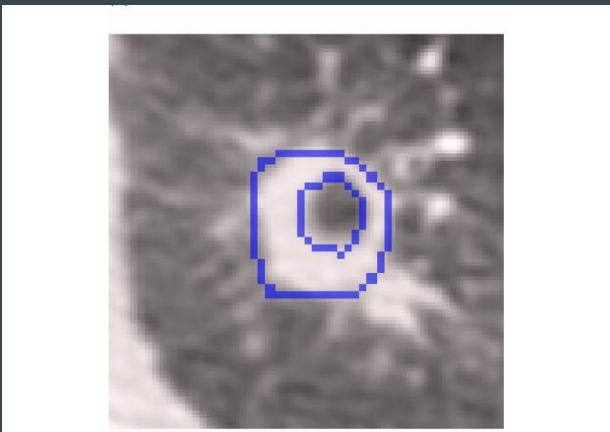
# LIDC expert review process

Each of the four radiologists independently reviewed all of the scans in a "blinded" phase to identify all of the findings from the three groups above.

For each finding identified by a given radiologist as a *"nodule ≥ 3 mm"*, outlines were constructed in each slice where that nodule appear, while for the other two categories only the approximate center of mass was annotated.

In the subsequent "unblinded" read phase each radiologist had access to the categories assigned and annotations for the nodules, and *"a radiologist's own marks then could be left unchanged, deleted, switched in terms of lesion category, or additional marks could be added"*.

After the unblinded phase each radiologist assessed subjective characteristics of "nodules < 3 mm", such as spiculation, subtlety, etc



solidity, margin, and likelihood of malignancy.[29] Each radiologist's lesion-category designation and associated marks (spatial locations of all points in the outlines constructed for a nodule ≥ 3 mm along with its characteristics and center-of-mass locations for a nodule < 3 mm and for a non-nodule ≥ 3 mm) for each lesion were stored in a single XML file for each scan after the unblinded read phase (the XML schema is located at http://troll.rad.med.umich.edu/lidc/). The blinded and unblinded read phases were intended to comprise a single, comprehensive process; therefore, the LIDC/IDRI Database only contains the final set of post-unblinded-read-phase marks in each of the 1018 XML files.

Armato et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **38**, 915–931 (2011).

```xml
<?xml version="1.0" encoding="UTF-8"?>
<LidcReadMessage uid="1.3.6.1.4.1.14519.5.2.1.6279.6001.1308168927827.0" xmlns="http://www.nih.gov">
 <ResponseHeader>
  <Version>1.7</Version>
  <MessageId>1151420</MessageId>
  <DateRequest>2006-06-05</DateRequest>
  <TimeRequest>17:05:26</TimeRequest>
  <RequestingSite>removed</RequestingSite>
  <ServicingSite>removed</ServicingSite>
  <TaskDescription>Second unblinded read</TaskDescription>
  <CtImageFile>removed</CtImageFile>
  <SeriesInstanceUid>1.3.6.1.4.1.14519.5.2.1.6279.6001.303494235102183795724852353824</SeriesInstanceUid>
  <StudyInstanceUID>1.3.6.1.4.1.14519.5.2.1.6279.6001.339170810277323131167631068432</StudyInstanceUID>
  <DateService>2006-06-05</DateService>
  <TimeService>17:05:26</TimeService>
  <ResponseDescription>1 - Reading complete</ResponseDescription>
  <ResponseComments>Merged, reader anonymized, unblinded responses</ResponseComments>
 </ResponseHeader>
```

```xml
<readingSession>
 <annotationVersion>3.12</annotationVersion>
 <servicingRadiologistID>anonymous</servicingRadiologistID>
 <unblindedReadNodule>
  <noduleID>0</noduleID>
  <roi>
   <imageZposition>1604.5</imageZposition>
   <imageSOP_UID>1.3.6.1.4.1.14519.5.2.1.6279.6001.192028152416081898151427003898</imageSOP_UID>
   <inclusion>TRUE</inclusion>
   <edgeMap><xCoord>177</xCoord><yCoord>353</yCoord></edgeMap>
  </roi>
 </unblindedReadNodule>
 <unblindedReadNodule>
  <noduleID>3</noduleID>
  <characteristics>
   <subtlety>5</subtlety>
   <internalStructure>1</internalStructure>
   <calcification>6</calcification>
   <sphericity>3</sphericity>
   <margin>4</margin>
   <lobulation>1</lobulation>
   <spiculation>1</spiculation>
   <texture>5</texture>
   <malignancy>3</malignancy>
  </characteristics>
```

# Consequences of using project-specific XML representation

- No single conventional tool that can render annotations over the image
- Annotations need to be managed separately from images
  - Accessed as a ZIP file
- Linked to images via DICOM Study/InstanceUID, patient information separate from the XML
- Conventions about stored content need to be parsed from the accompanying documentation and annotated XML example
- Conversion of contours into image-space is error-prone
  - inclusion/exclusion contours, inclusion or exclusion of boundary pixels, coordinate space conventions
- Representation not harmonized with any conventions for any other TCIA collection

# Tools/studies developed to enable interpretations of those XML files (*a subset, I am sure*)

## MAX

MAX ("multi-purpose application for XML") performs nodule matching and pmap generation based on the XML files provided with the LIDC/IDRI Database. It also performs certain QA and QC tasks and other XML-related tasks.

MAX is written in Perl and was developed under RedHat Linux. It has been run under Windows.

Downloading MAX and its associated files implies acceptance of the following notice (also available here and in the distro as a text file):

## LIDC 2 Image Toolbox (Matlab)

This tool is a community contribution developed by Thomas Lampert. It is designed for extracting individual annotations from the XML files and converting them, and the DICOM images, into TIF format for easier processing in Matlab (LIDC-IDRI dataset). It is available for download from: https://sites.google.com/site/tomalampert/code.

## pylidc

pylidc is an Object-relational mapping (using SQLAlchemy) for the data provided in the LIDC dataset. This means that the data can be queried in SQL-like fashion, and that the data are also objects that add additional functionality via functions that act on instances of data obtained by querying for particular attributes.

2297

## Development of a Data Integration and Visualization Software for LIDC

Chao Zeng[1, 2]
[1]School of Geosciences and Info-physics, Central South University, Changsha, China
[2]College of Information Science and Technology, Shihezi University, Shihezi, China
Email: zengc_bme@csu.edu.cn

Hongli Lin[*1, 3] and Weisheng Wang[3]
[*1]School of Geosciences and Info-physics, Central South University, Changsha, China
[3]College of Information Science and Engineering, Hunan University, Changsha, China
Email: hllin@hnu.edu.cn

ELSEVIER

Technical report

A Pulmonary Nodule View System for the Lung Image Database Consortium (LIDC)

Hongli Lin PhD [a, b], Zhencheng Chen PhD [c], Weisheng Wang MD [b]

# DICOM - preparing for the unknown, since 1983



- Standard for images <u>and image-related evidence</u>
- "The HOW": Fixed syntax, encoding, compression ...
  - (hierarchical) list of attribute/value pairs
- "The WHAT": Object definitions
  - Object-specific required and optional attributes
  - Constraints and values sets
  - Common data elements / lexicons / ontologies
- For all object types
  - Dates, patient IDs, study, series - for every object
  - Unique identifiers
- References to related evidence
  - Provenance of data acquisition, analysis
- + networking, web, de-identification ...



Radiologists and radiological physicists who were doing research using digital images were, at about the same time as the 1982 PACS meeting, complaining to the ACR about difficulties accessing these images. Since this involved medical imaging equipment that was regulated by the Food and Drug Administration's (FDA) Center for Devices and Radiological Health (at the time in transition from the Bureau of Radiological Health), the ACR made inquiries to the FDA about the problems the radiologists were having with digital images. The result was a meeting of representatives of equipment manufacturers (through NEMA), radiologists (through the ACR), and the FDA. The vendors agreed that a voluntary standard would be preferable to a regulatory one (SIIM, 2008). Shortly thereafter, in November 1983, the ACR and NEMA met to form the Digital Imaging and Communications Standards Committee (Horii, 2005).
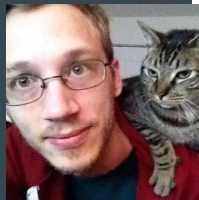
*Horii SC. DICOM. In: Kagadis GC, Langer SG, editors. Informatics in medical imaging. 2011. p. 41–67.*

# *pylidc*

"*pylidc is an Object-relational mapping (using SQLAlchemy) for the data provided in the LIDC dataset. This means that the data can be queried in SQL-like fashion, and that the data are also objects that add additional functionality via functions that act on instances of data obtained by querying for particular attributes*"

- Open source, python
- Easy to install on any platform
- Used by others!
- More than LIDC parser
- Supported
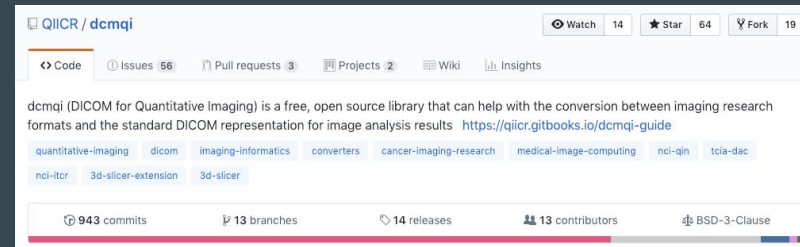  - Thanks Matt Hancock!



Hancock, M. C. & Magnan, J. F. Lung nodule malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms: probing the Lung Image Database Consortium dataset with two statistical learning methods. *J Med Imaging (Bellingham)* **3,** 044504 (2016).

# *dcmqi*      [https://github.com/qiicr/dcmqi](https://github.com/qiicr/dcmqi)

"dcmqi (DICOM for Quantitative Imaging) is a free, open source library that can help with the conversion between imaging research formats and the standard DICOM representation for image analysis results"

- DICOM Segmentation object
- DICOM Parametric map object
- DICOM Region-based measurements Structured Report

Herz, C., Fillion-Robin, J.-C., Onken, M., Riesmeier, J., Lasso, A., Pinter, C., Fichtinger, G., Pieper, S., Clunie, D., Kikinis, R. & Fedorov, A. dcmqi: An Open Source Library for Standardized Communication of Quantitative Image Analysis Results Using DICOM. *Cancer Res.* **77,** e87–e90 (2017).

# LIDC nodule segmentations

- (x,y) coordinates of the contour points for image slices defined by DICOM SOPInstanceUID
- Contours can be "inclusion" or exclusion
- Contour points are not part of the segmentation!
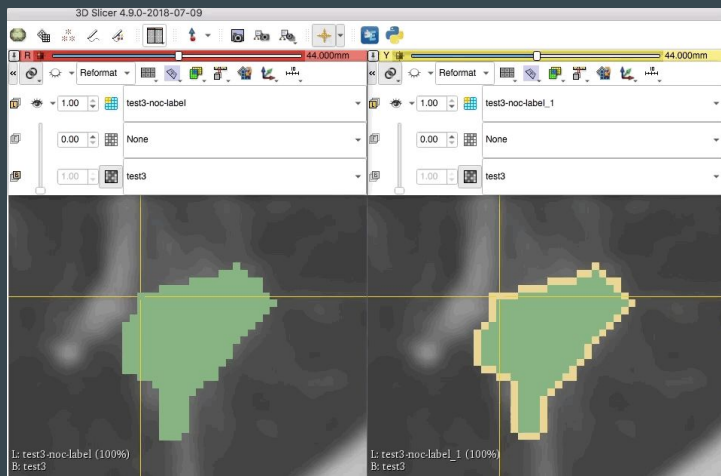


4. Nodule Contour ROI – this is the description of the complete three dimensional contour of the nodule (remembering that the radiologist was instructed to mark the first voxel outside the nodule).

Things to NOTE:

The overall format is to report each z (longitudinal direction) position using ImageZposition and imageSOP UID below) on which the nodule is visualized; and within each z position, to report the x and y coordinates (Note that (0,0) is upper left of the image) of the connected boundary points within each x-y plane.

Note also that the LIDC allowed the radiologists to describe regions of exclusion (primarily regions of air within a nodule), so an "inclusion" tag was developed: when the "inclusion" value is "true" the roi being described is considered to be part of the nodule, when the "inclusion" value is "false, the roi being described is excluded (or subtracted from) the nodule.

https://github.com/notmatthancock/pylidc/issues/15          https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

# Image Segmentations - as DICOM

Spatial labeling of the voxels of interest +

- Patient information
- Unique identifiers
- References to the segmented image
- Ontology-enabled semantics (SNOMED-CT, RadLex, etc)
- Segmentation algorithm
- Single file, efficient storage, multiple occupancy voxels
- Presentation color

| SeriesNumber | SeriesDate | SeriesTime | SeriesDescription | Modality | BodyPartExamined |
|---|---|---|---|---|---|
| 0 | 1988-09-13 | 102727.421000 | CT HeadNeck  3.0  B30f_CHEST | CT | HEADNECK |
| 0 | 1988-09-13 | 102731.687000 | PET HeadNeck_0 | PT | HEADNECK |
| 1117 | 2015-08-19 | 121219 | PET SUV Factors | RWV | HEADNECK |
| 1146 | 2015-08-19 | 122027 | tumor segmentation – User3 SemiAuto trial 2 | SEG | HEADNECK |
| 1144 | 2015-08-19 | 121944 | tumor segmentation – User3 SemiAuto trial 1 | SEG | HEADNECK |
| 1142 | 2015-08-19 | 121900 | tumor segmentation – User3 Manual trial 2 | SEG | HEADNECK |
| 1140 | 2015-08-19 | 121818 | tumor segmentation – User3 Manual trial 1 | SEG | HEADNECK |
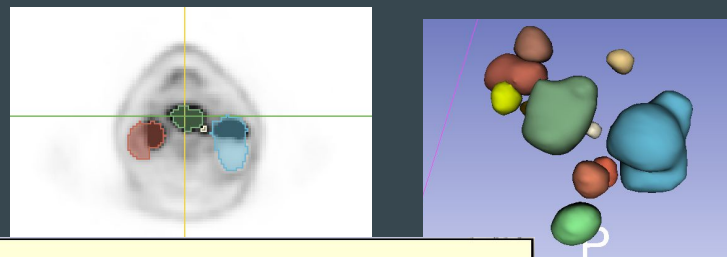
Terminology:
Context: PET HeadNeck_0 tumor segmentation - User3 SemiAuto trial 2 - Segmentations
Category: Morphologically Altered Structure
Type: Neoplasm, Primary
Anatomic context: PET HeadNeck_0 tumor segmentation - User3 SemiAuto trial 2 - Segmentations
Anatomic region: base of tongue

```
(0008,0018) UI [1.2.276.0.7230010.3.1.4.8323329.17023.1440004576.837084] #  56, 1 SOPInstanceUID
(0008,0020) DA [19880913]                                                #   8, 1 StudyDate
(0008,0021) DA [20150819]                                                #   8, 1 SeriesDate
(0008,0023) DA [20150819]                                                #   8, 1 ContentDate
(0008,0030) TM [101844.421000]                                           #  14, 1 StudyTime
(0008,0031) TM [121617]                                                  #   6, 1 SeriesTime
(0008,0033) TM [121617]                                                  #   6, 1 ContentTime
(0008,0050) SH [2076699673350889]                                        #  16, 1 AccessionNumber
(0008,0060) CS [SEG]                                                     #   4, 1 Modality
(0008,0070) LO [QIICR]                                                   #   6, 1 Manufacturer
(0008,0090) PN (no value available)                                      #   0, 0 ReferringPhysicianName
(0008,1030) LO [CT CHEST W/O CONTRAST]                                   #  22, 1 StudyDescription
```

# LIDC segmentations: conversion process

- One nodule annotation per DICOM Segmentation object
- Use *pylidc* to get numpy array for each nodule segmentation
  - Also use pylidc clustering functionality to assign segmentations to individual nodules
- Use *ITK-python* to reorient the array, add origin/spacing/directions and save as a 3D volume
- Populate metadata JSON
- Use *dcmqi* to write DICOM Segmentation object
- Conversion process includes pointer to the source CT series
  - Propagate composite context
  - Establish references to the source series analyzed

https://itkpythonpackage.readthedocs.io

```
"segmentAttributes": [
  [
   {
    "SegmentLabel": "Nodule 1 - Annotation 157143",
    "recommendedDisplayRGBValue": [
     241,
     214,
     145
    ],
    "AnatomicRegionSequence": {
     "CodeMeaning": "Lung","CodingSchemeDesignator": "SRT","CodeValue": "T-28000"
    },
    "SegmentAlgorithmType": "MANUAL",
    "SegmentedPropertyCategoryCodeSequence": {
     "CodeMeaning": "Morphological Abnormal Structure","CodingSchemeDesignator": "SRT","CodeValue":
"M-01000"
    },
    "SegmentedPropertyTypeCodeSequence": {
     "CodeMeaning": "Nodule","CodingSchemeDesignator": "SRT","CodeValue": "M-03010"
    },
    "TrackingIdentifier": "Nodule 1",
    "labelID": 1,
    "SegmentDescription": "Nodule 1 - Annotation 157143",
    "TrackingUniqueIdentifier": "2.25.78955135423773784954011632177062973466658992040550559735258"
   }
  ]
]
```

# LIDC nodule characterization

```xml
<characteristics>
  <subtlety>5</subtlety>
  <internalStructure>1</internalStructure>
  <calcification>6</calcification>
  <sphericity>3</sphericity>
  <margin>4</margin>
  <lobulation>1</lobulation>
  <spiculation>1</spiculation>
  <texture>5</texture>
  <malignancy>3</malignancy>
</characteristics>
```

```
<characteristics>
        For a nodule > 3 mm, each reader is asked to subjectively assess
the nodule's characteristics as described on LIDC CDE page
http://cdebrowser.nci.nih.gov/CDEBrowser/ and as follows:
        <subtlety>5</subtlety>
        Radiologist assessment of nodule subtlety on 1-5 scale
(1      extremely subtle
2
3
4
5       obvious)
        <internalStructure>1</internalStructure>
        Radiologist assessment of nodule internal structure
(1      soft tissue,
2       fluid,
3       fat,
4       air)
        <calcification>6</calcification>
        Radiologist assessment of internal calcification of nodule
(1      Popcorn         Popcorn Appearance
2       Laminated       Laminated Appearance
3       Solid           Solid Appearance
4       Non-Central     Non-Central Appearance
5       Central         Central Calcification
6       Absent)
        <sphericity>4</sphericity>
 Radiologist assessment of shape of nodule in terms of its
roundness/sphericity with only 3 terms defined:
(1      Linear          Linear Appearance
2
3       Ovoid           Ovoid Appearance
4
5       Round           Round Appearance)
        <margin>4</margin>
```

# DICOM Structured Reporting (SR)

- Extends the DICOM standard beyond attributes in "image header"
- Into a "tree" of structured content
- Each tree node ("content item") is a "name-value" pair
- Or a "container" (provides an outline of "headings")
- Name is always coded, value may or may not be
- Generic framework - recursively nested content of unbounded complexity
- Complexity constrained for specific use cases by "templates"
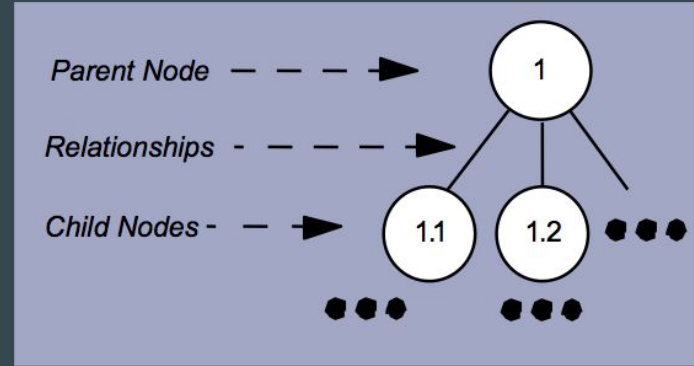
# DICOM SR TID1500: measurements + characterizations

- TID 1500 "Measurement report"
- General-purpose template
- Reuses existing templates
- Planar or volumetric region definition
- Qualitative characterizations

Example from *Fedorov et al. DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. PeerJ 4, e2057 (2016).*
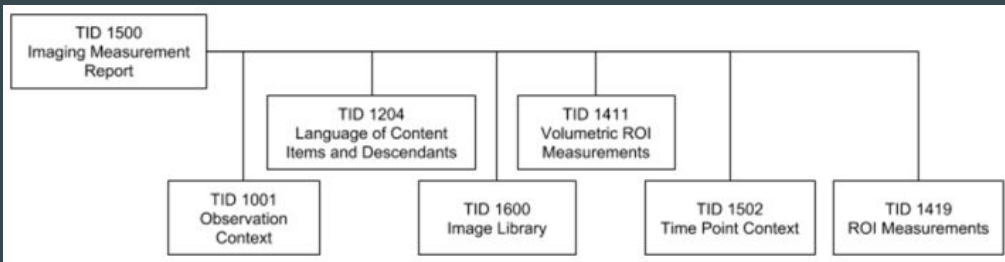
# DICOM SR approach to coded content

- Qualitative assessment reporting
  - Coded concept
  - Coded value
- Coded = (Code Value, Coding Scheme Designator, Code Meaning) tuple
  - E.g., ("CodeMeaning": "Lung","CodingSchemeDesignator": "SRT","CodeValue": "T-28000")
- Sources of codes
  - Standard dictionaries: SNOMED, NCIt, RadLex, UCUM, …
  - "Private" dictionaries: when standard dictionaries do not contian  suitable terms, indicated by "99" prefix of the Coding Scheme Designator

# Harmonization of LIDC with existing terminologies

- Review existing lexicons: NCI Thesaurus, RadLex, SNOMED (only those terms included in the DICOM standard)
  - BioPortal http://bioportal.bioontology.org/
  - Ontology Lookup Service https://www.ebi.ac.uk/ols
- Imaging Biomarker Standardization Initiative (IBSI) lexicon
- "99LIDCQIICR" coding scheme for codes not located in existing lexicons

Zwanenburg, A., Leger, S., Vallières, M., Löck, S. & for the Image Biomarker Standardisation Initiative. Image biomarker standardisation initiative. *arXiv [cs.CV]* (2016). at http://arxiv.org/abs/1612.07003

**Table 1. LIDC-IDRI Evaluation Concepts**

| Coding Scheme Designator | Code Value | Code Meaning | LIDC-IDRI concept name verbatim | LIDC definition |
|---|---|---|---|---|
| NCIt | C45992 | Subtlety Score | Subtlety | Radiologist assessment of nodule subtlety on 1-5 scale |
| 99LIDCQIICR | 200 | Internal structure | Internal structure | Radiologist assessment of nodule internal structure |
| NCIt | C3672 | Calcification | Calcification | Radiologist assessment of internal calcification of nodule |
| IBSI | QCFX | Sphericity | Sphericity | Radiologist assessment of shape of nodule in terms of its roundness/sphericity with only 3 terms defined |
| NCIt | C25563 | Margin | Margin | Radiologist assessment of |

# Coding concept values

- Some concept values are categorical (e.g., "Internal structure")
- Others are ordinal scores
  - With or without specific assignment of meaning

| Margin | | | | |
|---|---|---|---|---|
| RadLex | RID5709 | Indistinct margin (synonym: poorly defined margin) | 1 - Poorly defined |
| 99LIDCQIICR | 002 | 2 out of 5 | 2 |
| 99LIDCQIICR | 003 | 3 out of 5 | 3 |
| 99LIDCQIICR | 004 | 4 out of 5 | 4 |
| RadLex | RID5707 | Circumscribed margin (synonym: | 5 - Sharp margin |

| Internal structure | | | | |
|---|---|---|---|---|
| NCIt | C12471 | Soft tissue | Soft tissue |
| NCIt | C25278 | Fluid | Fluid |
| NCIt | C12472 | Adipose tissue | Fat |
| NCIt | C73434 | Air | Air |

# Results of mapping

- Concepts (9 total):
  - 6 mapped to NCIt
  - 1 to RadLex
  - 1 to IBSI
  - 1 not mapped (99LIDCQIICR)
- Values (45 total):
  - 29 not mapped (99LIDCQIICR)
  - 12 to RadLex
  - 4 to NCIt

Authors          Authors and affiliations

Pia Opulencia  1        1.   School of Computing, DePaul University, Chicago, USA
  Email author
                        2.   Department of Radiology, Feinberg School of Medicine, Northwestern
David S. Channin  2          University, Chicago, USA

Daniela S. Raicu  1

Jacob D. Furst  1

*Coordinated with the Opulencia et al. work after v1!*

```
<characteristics>
  <subtlety>5</subtlety>
  <internalStructure>1</internalStructure>
  <calcification>6</calcification>
  <sphericity>2</sphericity>
  <margin>3</margin>
  <lobulation>3</lobulation>
  <spiculation>3</spiculation>
  <texture>4</texture>
  <malignancy>3</malignancy>
</characteristics>
```

```
C3672,NCIt,"Calcification")=(RID28473,RadLex,"Absent")>
(200,99LIDCQIICR,"Internal structure")=(C12471,NCIt,"Soft
tissue")>
(400,99LIDCQIICR,"Sphericity")=(002,99LIDCQIICR,"2 out of 5")>
(C45992,NCIt,"Subtlety score")=(105,99LIDCQIICR,"5 out of 5
(Obvious)")>
(700,99LIDCQIICR,"Spiculation")=(003,99LIDCQIICR,"3 out of 5")>
(600,99LIDCQIICR,"Lobulation")=(003,99LIDCQIICR,"3 out of 5")>
(C25563,NCIt,"Margin")=(003,99LIDCQIICR,"3 out of 5")>
(C41144,NCIt,"Texture")=(004,99LIDCQIICR,"4 out of 5")>
(900,99LIDCQIICR,"Malignancy")=(903,99LIDCQIICR,"3 out of 5
(Indeterminate Likelihood)")>
```

# DICOM SR conversion process

- One nodule characterization per DICOM Segmentation object
- Use *pylidc* to get characterizations for each nodule
  - Also use pylidc to get nodule volume, surface area, 2D diameter
- Populate metadata JSON
- Use *dcmqi* to write DICOM SR object
- Conversion process includes pointer to the source CT series and DICOM Segmentation object
  - Propagate composite context
  - Establish references to the source series analyzed
  - Establish references to the segmentation associated with the characterizations

```
(0008,0012) DA [20181119]                                      #   8, 1 InstanceCreationDate
(0008,0013) TM [155338]                                        #   6, 1 InstanceCreationTime
(0008,0014) UI [1.2.276.0.7230010.3.0.3.6.3]                   #  28, 1 InstanceCreatorUID
(0008,0016) UI =EnhancedSRStorage                              #  30, 1 SOPClassUID
(0008,0018) UI [1.2.276.0.7230010.3.1.4.0.49881.1542660818.290145] #  50, 1 SOPInstanceUID
(0008,0020) DA [20000101]                                      #   8, 1 StudyDate
(0008,0021) DA [20181119]                                      #   8, 1 SeriesDate
(0008,0023) DA [20181119]                                      #   8, 1 ContentDate
(0008,0030) TM (no value available)                            #   0, 0 StudyTime
(0008,0031) TM [155338]                                        #   6, 1 SeriesTime
(0008,0033) TM [155338]                                        #   6, 1 ContentTime
(0008,0050) SH [2819497684894126]                              #  16, 1 AccessionNumber
(0008,0060) CS [SR]                                            #   2, 1 Modality
(0008,0070) LO [QIICR]                                         #   6, 1 Manufacturer
(0008,0090) PN (no value available)                            #   0, 0 ReferringPhysicianName
(0008,103e) LO [Nodule 7 – Annotation IL057_167525 evaluations] #  46, 1 SeriesDescription
(0008,1090) LO [https://github.com/QIICR/dcmqi.git]            #  34, 1 ManufacturerModelName
(0008,1111) SQ (Sequence with undefined length #=0)            # u/l, 1 ReferencedPerformedProcedureStepSequence
(fffe,e0dd) na (SequenceDelimitationItem)                      #   0, 0 SequenceDelimitationItem
(0010,0010) PN (no value available)                            #   0, 0 PatientName
(0010,0020) LO [LIDC–IDRI–1005]                                #  14, 1 PatientID
(0010,0030) DA (no value available)                            #   0, 0 PatientBirthDate
(0010,0040) CS (no value available)                            #   0, 0 PatientSex
(0018,1000) LO [0]                                             #   2, 1 DeviceSerialNumber
(0018,1020) LO [ef688c4]                                       #   8, 1 SoftwareVersions
(0020,000d) UI [1.3.6.1.4.1.14519.5.2.1.6279.6001.560355181698089203099323359955] #  64, 1 StudyInstanceUID
(0020,000e) UI [1.2.276.0.7230010.3.1.3.0.49881.1542660818.290147] #  50, 1 SeriesInstanceUID
(0020,0010) SH (no value available)                            #   0, 0 StudyID
(0020,0011) IS [3000685]                                       #   8, 1 SeriesNumber
(0020,0013) IS [1]                                             #   2, 1 InstanceNumber
```

: CONTAINER: Imaging Measurement Report [SEPARATE] (DCMR,1500)
　HAS CONCEPT MOD: CODE: Language of Content Item and Descendants = English
　HAS OBS CONTEXT: CODE: Observer Type = Person
　HAS OBS CONTEXT: PNAME: Person Observer Name = anonymous
　HAS CONCEPT MOD: CODE: Procedure reported = Imaging procedure
▶ CONTAINS: CONTAINER: Image Library [SEPARATE] (DCMR,1600)
▼ CONTAINS: CONTAINER: Imaging Measurements [SEPARATE]
　　▼ CONTAINS: CONTAINER: Measurement Group [SEPARATE] (DCMR,1411)
　　　　HAS OBS CONTEXT: TEXT: Activity Session = 1
　　　　HAS OBS CONTEXT: TEXT: Tracking Identifier = Nodule 7
　　　　HAS OBS CONTEXT: UIDREF: Tracking Unique Identifier = 2.25.129840025507180422315945787493725797361488077138541131291247
　　　　CONTAINS: CODE: Finding = Nodule
　　　　HAS OBS CONTEXT: TEXT: Time Point = 1
　　　　CONTAINS: IMAGE: Referenced Segment = 1.2.840.10008.5.1.4.1.1.66.4 : 1.2.276.0.7230010.3.1.4.0.49880.1542660811.119023[Segment 1]
　　　　CONTAINS: UIDREF: Source series for segmentation = 1.3.6.1.4.1.14519.5.2.1.6279.6001.142485715518010940961688015191
　　　　HAS CONCEPT MOD: CODE: Finding Site = Lung
　　▶ CONTAINS: NUM: Volume = 4.123154E+01 cubic millimeter
　　▶ CONTAINS: NUM: Diameter = 6.299850E+00 millimeter
　　▶ CONTAINS: NUM: Surface area of mesh = 5.569699E+01 square millimeter
　　　　CONTAINS: CODE: Calcification = Absent
　　　　CONTAINS: CODE: Internal structure = Soft tissue
　　　　CONTAINS: CODE: Sphericity = 4 out of 5
　　　　CONTAINS: CODE: Subtlety score = 1 out of 5 (Extremely subtle)
　　　　CONTAINS: CODE: Spiculation = 1 out of 5 (No spiculation)
　　　　CONTAINS: CODE: Lobulation = 1 out of 5 (No lobulation)
　　　　CONTAINS: CODE: Margin = 2 out of 5
　　　　CONTAINS: CODE: Texture = non-solid pulmonary nodule
　　　　CONTAINS: CODE: Malignancy = 3 out of 5 (Indeterminate Likelihood)

# Status

- Conversion results available on TCIA since RSNA 2018 (Google Drive folder)
- Conversion pipeline script is available on GitHub
- Details in preprint (v1, **will** be updated!) - <u>feedback is welcomed!</u>

# LIDC DICOM in context: Support in tools

- DICOM Segmentation object: supported in multiple tools (see https://dicom4qi.readthedocs.io/)
    - Open source: 3D Slicer, ePAD, MITK, MeVisLab, OHIF Viewer + dcmjs
    - Commercial: Brainlab, Siemens syngo.via, Terarecon (more in development)
- DICOM TID1500 SR:
    - 3D Slicer, ePAD can read and write
    - Pixelmed tools
    - Used in RSNA Crowds Cure Cancer for planar annotations+measurements
    - General-purpose DCMTK tools can be used to convert to XML representation (both atribute-level and SR tree level)
- Conversion of measurements into tabulated form
    - https://github.com/QIICR/dcm2tables

# Major DICOM4QI participating tools



Academic

Commercial

Production
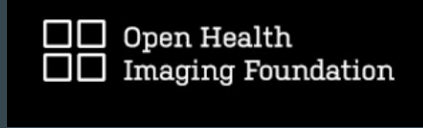
Prototype/under development

# LIDC DICOM in context: harmonization of TCIA derived data

- Numerous other collections follow "the DICOM approach" for derived data
  - **QIN-HEADNECK**: head and neck cancer, PET/CT, SUV, segmentations, SUV quantification
  - **QIN-PROSTATE-Repeatability**: prostate cancer, multiparametric MRI, segmentations, measurements
  - **TCIA-LGG and TCIA-GBM**: glioblastoma, multiparametric MRI, segmentations
- Extensible with other types of derived data
  - Framework for analysis from multiple groups/tools to be combined
  - pyradiomics supports output of radiomics features as DICOM SR TID1500
  - IBSI lexicon of radiomics features is being integrated with the DICOM Standard



✓ PEER-REVIEWED Bioinformatics and Genomics section ›

DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research

Research article | Bioinformatics | Clinical Trials | Oncology | Radiology and Medical Imaging

Andriy Fedorov [1,2], David Clunie [3], Ethan Ulrich [4,5], Christian Bauer [4,5], Andreas Wahle [4,5], Bartley Brown [6], Michael Onken [7], Jörg Riesmeier [8], Steve Pieper [9], Ron Kikinis [1,2,10,11], John Buatti [12], Reinhard R. Beichel [4,5,13]

Published May 24, 2016

https://peerj.com/articles/2057/



MENU ⌄    SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 04 December 2018

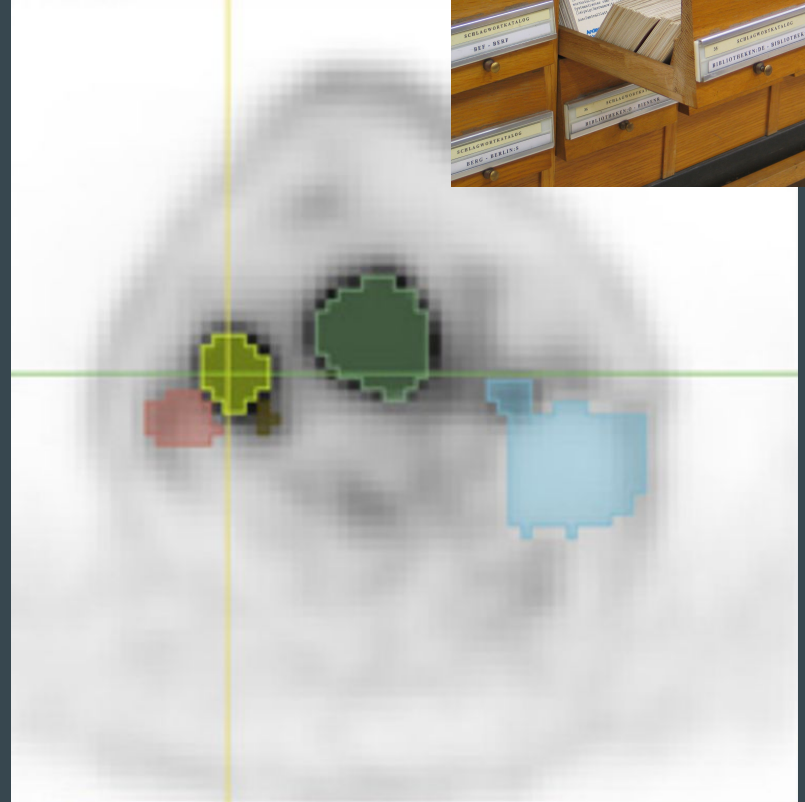An annotated test-retest collection of prostate multiparametric MRI

Andriy Fedorov, Michael Schwier, David Clunie, Christian Herz, Steve Pieper, Ron Kikinis, Clare Tempany & Fiona Fennessy

Scientific Data **5**, Article number: 180281 (2018) | Download Citation ⌄

https://www.nature.com/articles/sdata2018281

# Queries like this become possible:



Find all subjects that have a neoplasm segmented using automatic method at the base of tongue larger than 5 cc that show Glycolysis Within First Quarter of Intensity Range above 10 grams

# Acknowledgments



http://qiicr.org