# RCC52
# Harmonizing TCIA image-related data using DICOM

•••

Andrey Fedorov, PhD
Brigham and Women's Hospital, Boston
andrey.fedorov@gmail.com

29 November 2018

# *ImageNet* and *Large Scale Visual Recognition Challenge*

**Summary and Statistics (updated on April 30, 2010)**

**Overall**

- Total number of non-empty synsets: 21841
- Total number of images: 14,197,122
- Number of images with bounding box annotations: 1,034,908
- Number of synsets with SIFT features: 1000
- Number of images with SIFT features: 1.2 million

**Knowledge structure** [ edit ]

Both nouns and verbs are organized into hierarchies, defined by hypernym or *IS A* relationships. For instance, one sense of the word *dog* is found following hypernym hierarchy; the words at the same level represent synset members. Each set of synonyms has a unique index.

dog, domestic dog, Canis familiaris
- canine, canid
  - carnivore
    - placental, placental mammal, eutherian, eutherian mammal
      - mammal
        - vertebrate, craniate
          - chordate
            - animal, animate being, beast, brute, creature, fauna
              - ...

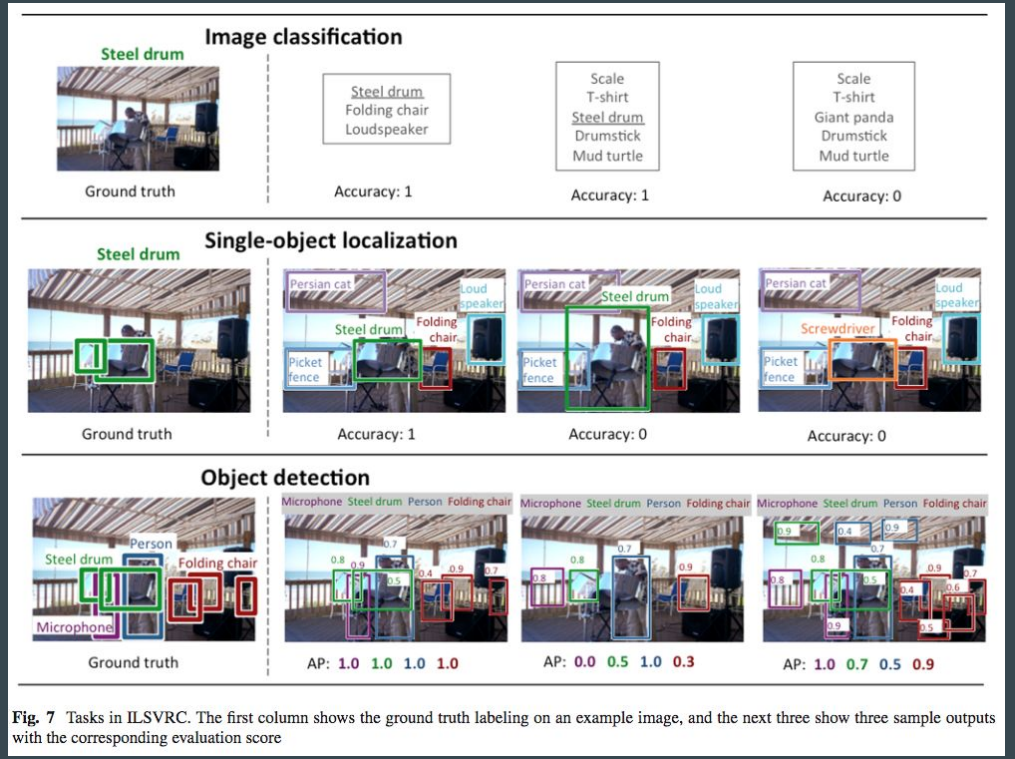


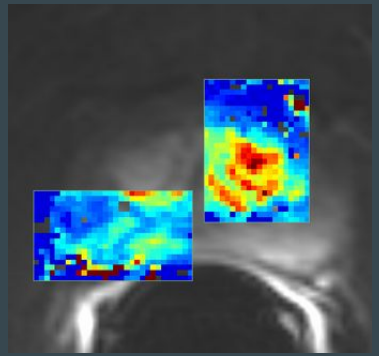

**Fig. 7** Tasks in ILSVRC. The first column shows the ground truth labeling on an example image, and the next three show three sample outputs with the corresponding evaluation score

http://image-net.org/challenges/LSVRC/2017/index
http://www.image-net.org/about-stats
https://wordnet.princeton.edu/

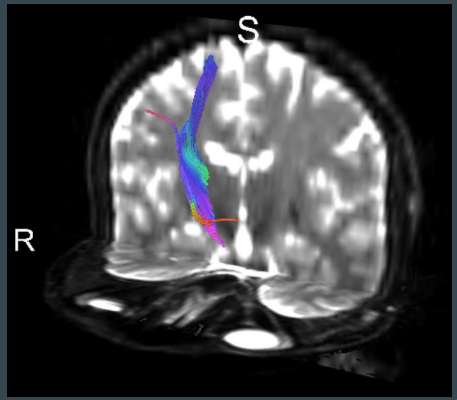
*Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis. Springer US; 2015;115(3):211–252*
*http://www.image-net.org/*

# Image-related data in medical imaging

- Image-like (e.g., segmentations, parametric maps)
- Non-image-like
  - Quantitative, qualitative, categorical (e.g., measurements, impressions)
  - White matter tracks, centerlines, planar annotations, …
- Clinical data (labs, therapies, outcomes)



| | Segment | Mean [{SUVbw}g/ml] | Minimum [{SUVbw}g/ml] |
|---|---|---|---|
| 1 | primary tumor | 10.6175 | 5.0406 |
| 2 | lymph node 1 | 5.18311 | 3.76263 |
| 3 | lymph node 2 | 8.29498 | 3.88878 |
| 4 | lymph node 3 | 3.38146 | 0.353775 |
| 5 | lymph node 4 | 6.27497 | 3.60357 |
| 6 | lymph node 5 | 8.73281 | 4.22884 |
| 7 | lymph node 6 | 8.50875 | 3.00571 |



# FAIR data: Findable Accessible Interoperable Reusable

**Box 2** | The FAIR Guiding Principles

**To be Findable:**
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016).

# Data and Metadata


"Metadata is a map. Metadata is a means by which the complexity of an object is represented in the simplest form."

"Without the data about objects contained in a space, any sufficiently complex space is indistinguishable from chaos"

Pomerantz, J. *Metadata (The MIT Press Essential Knowledge series)*. (The MIT Press, 2015).

# DICOM - preparing for the unknown, since 1983

- Standard for images and image-related evidence
- "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 …

DICOM℠
Digital Imaging and Communications in Medicine

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.*

# Repeatability of mpMRI

- Confirmed or suspected PCa
- Quantitative measures:
  - Expert annotations
  - Image-based volume and mean Apparent Diffusion Coefficient
- Repeat mpMRI within 2 weeks, with e-coil
- 189 men approached, 40 consented, **15 completed the study**
- Small volume disease: MRI-estimated tumor volume was less than 0.5 cc in 8 of 11 subjects

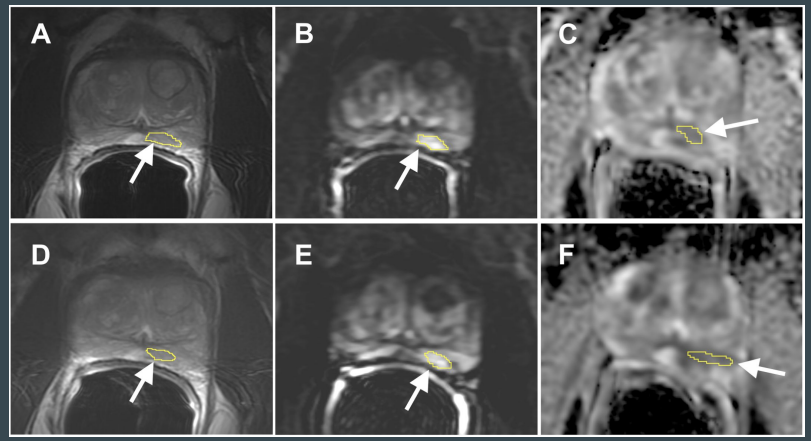



***Table 1*** *Repeatability of the region of interest volume measurements for different structures segmented on T2-weighted axial images, and ADC and SUB maps.*

| | RC% | RC, cc | Mean difference, cc (% mean difference) | ICC |
|---|---|---|---|---|
| ***PZ tROI*** | | | | |
| ADC | 112.2 | 0.4 | 0.1(42.4%) | 0.7 |
| SUB | 119.4 | 0.4 | 0.1 (42.8%) | 0.57 |
| T2AX | 70.5 | 0.2 | 0.08 (25.5%) | 0.86 |

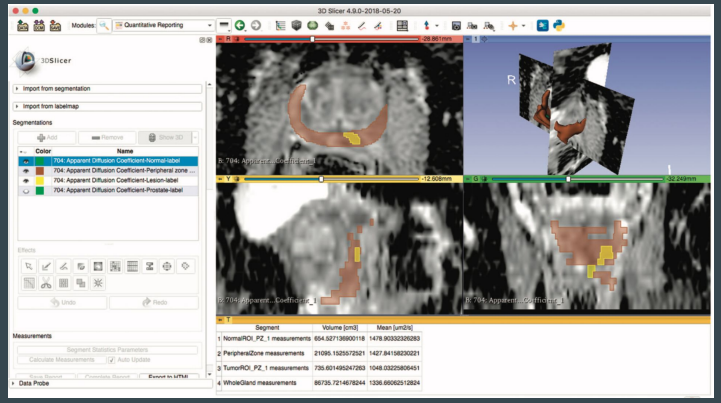
***Table 2*** *Repeatability of the mean ADC measurements (b0-1400) for the segmented structures.*

| | RC% | RC, $\times 10^{-6}$ $mm^{-2}/sec$ | Mean difference, $\times 10^{-6}$ $mm^{-2}/sec$ (% mean difference) | ICC |
|---|---|---|---|---|
| WG | 29.5 | 359 | 83 (6.85%) | 0.72 |
| PZ | 22.3 | 305 | 88 (6.45%) | 0.68 |
| nPZ | 30.2 | 471 | 175 (11.27%) | 0.46 |
| PZ tROI | 41.8 | 447 | 170 (15.93%) | 0.3 |

Fedorov, A., Vangel, M. G., Tempany, C. M. & Fennessy, F. M. Multiparametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. *Invest. Radiol.* **52,** 538–546 (2017).

# TCIA QIN-Prostate-Repeatability collection

- Multiparametric MRI at 2 time points: T2w, ADC, DCE subtract series
- Expert annotations of the regions of interest (prostate gland, peripheral zone, suspected tumor, normal tissue)
- Volume and mean ADC for the annotated regions
- TCIA limited access collection




SCIENTIFIC DATA

OPEN

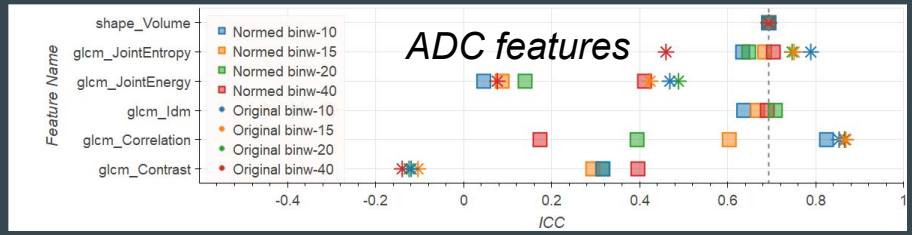
Data Descriptor: An annotated test-retest collection of prostate multiparametric MRI

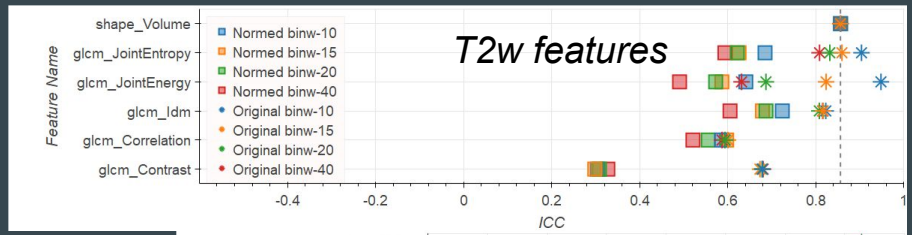
Andriy Fedorov[1], Michael Schwier[1], David Clunie[2], Christian Herz[1], Steve Pieper[3], Ron Kikinis[1,4,5], Clare Tempany[1] & Fiona Fennessy[1]

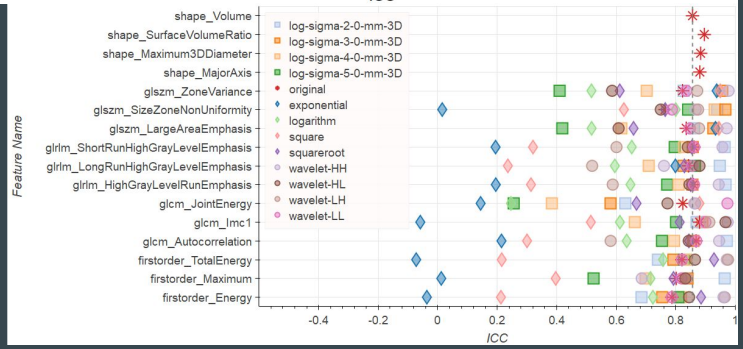
Fedorov, A., Schwier, M., Clunie, D., Herz, C., Pieper, S., Kikinis, R., Tempany, C. & Fennessy, F. An annotated test-retest collection of prostate multiparametric MRI. *Scientific Data* **5,** 180281 (2018) https://www.nature.com/articles/sdata2018281

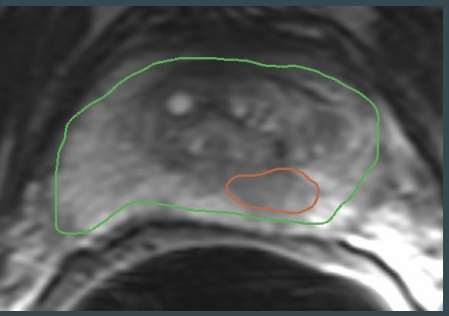
# Repeatability of mpMRI Radiomics features in the prostate

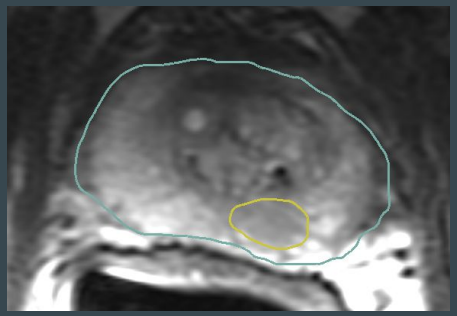
- *pyradiomics* for feature extraction
- Extraction be highly sensitive to extraction parameters
- Features identified as predictive may not be reproducible (and *vice versa*!)





*Schwier M, van Griethuysen J, Vangel MG, Pieper S, Peled S, Tempany CM, Aerts H, Kikinis R, Fennessy FM, Fedorov A. Repeatability of Multiparametric Prostate MRI Radiomics Features. arXiv [cs.CV] (2018). https://arxiv.org/abs/1807.06089*

# Optimizing DCE MRI analysis in the prostate

- Analysis covariates explored: Arterial Input Function, Pharmacokinetic model, Bolus arrival time
- Reflux rate constant ($k_{ep}$) more repeatable than $K^{trans}$
- No difference in discriminating between suspected cancer and normal tissue
- DCE MRI series, PK maps will be added to the TCIA-Prostate-Repeatability collection

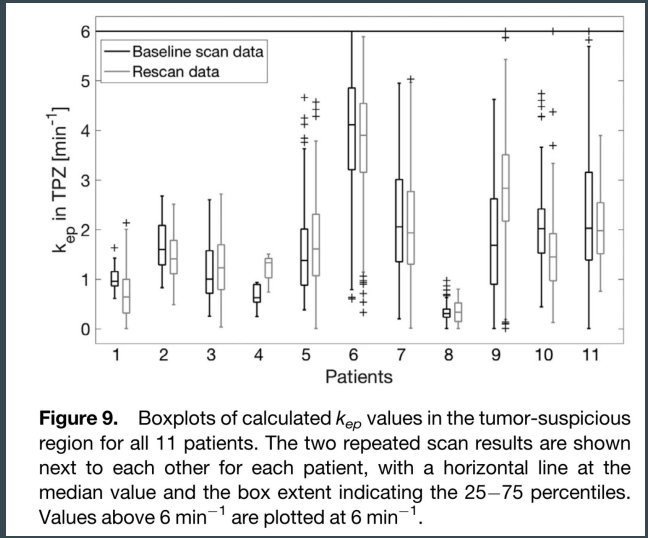
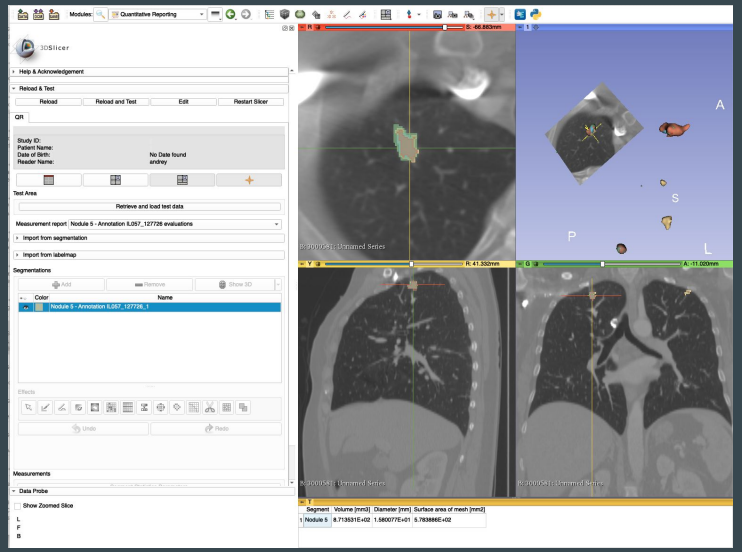



**Figure 9.** Boxplots of calculated $k_{ep}$ values in the tumor-suspicious region for all 11 patients. The two repeated scan results are shown next to each other for each patient, with a horizontal line at the median value and the box extent indicating the 25−75 percentiles. Values above 6 min$^{-1}$ are plotted at 6 min$^{-1}$.


Peled, S., Vangel, M., Kikinis, R., Tempany, C. M., Fennessy, F. M. & Fedorov, A. Selection of Fitting Model and Arterial Input Function for Repeatability in Dynamic Contrast-Enhanced Prostate MRI. *Acad. Radiol.* (2018). https://doi.org/10.1016/j.acra.2018.10.018

# TCIA LIDC-IDRI annotations

- Lung Image Database Consortium (LIDC): Muti-site data acquisition and annotation project
- Chest CT for 1000+ subjects, localization and characterization of nodules ≥3 by 4 experts
- Volumetric annotation of nodules
- Qualitative assessment score for margin, spiculation, calcification, etc.
- Annotations originally stored using project-specific XML representation






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).

Fedorov, A., Hancock, M., Clunie, D., Brochhausen, M., Bona, J., Kirby, J., Freymann, J., Pieper, S., Aerts, H., Kikinis, R. & Prior, F. *Standardized representation of the LIDC annotations using DICOM*. (PeerJ Preprints, 2018). https://doi.org/10.7287/peerj.preprints.27378

# DICOM for image-derived data means:

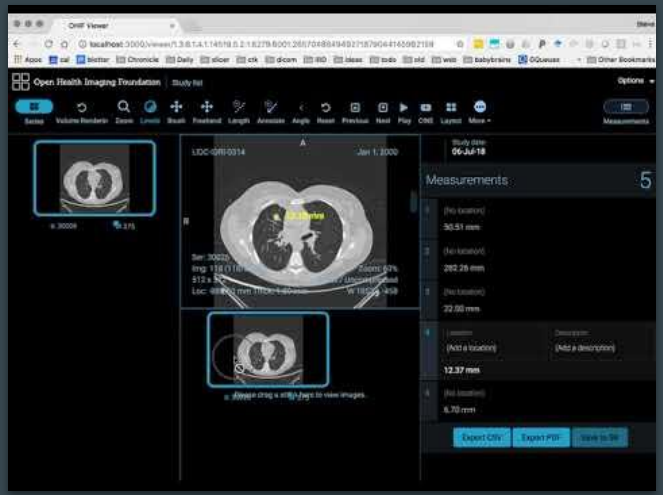
- For each object: subject identification, dates, unique identifiers, object type
- Segmentations: SNOMED semantics for segmented structure, references to segmented images, segmentation method identification
- Measurements (DICOM SR): coded semantics for quantities and units, references to segmentations and images, identification of the algorithm and parameters
- Uniform approach for encoding derived data across collections
  - Aggregate queries of image-derived data across collections become possible
- Generic tools applicable for visualization/query/conversion
- Same database for images and image-derived data
- Pathway for using clinical workstations for generating analysis results

# Examples of queries that now become possible

- *“Find all female subjects that have nodules located in the lung, and which are larger than 1 cc”*
- *“Find all structures segmented by more than one algorithm”*
- *“Find all subjects that had imaging at more than one time point, with secondary tumor in the neck region”*
- *“Plot correlation between the expert-assigned spiculation score and calculated sphericity for all segmented lung nodules”*

# Open source tools

- dcmqi: library and command line tools for conversion
- pyradiomics: can generate radiomics featues stored as DICOM SR
- Variety of off-the-shelf tools suitable for consuming standard representation:
  - Desktop: 3D Slicer, MITK, MeVisLab, …
  - Web: ePAD, OHIF Viewer and dcmjs (work in progress)
- DCMTK: conversion into JSON and XML
- Tools for generating tabular view

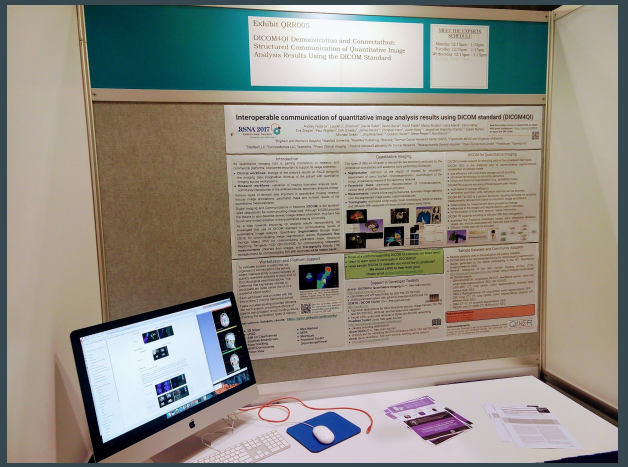




https://github.com/qiicr/dcmqi

# Promoting and evaluating adoption of the standard

*DICOM4QI: DICOM for Quantitative Imaging* https://dicom4qi.readthedocs.io/

- Demonstration and connectathon at RSNA QIRR since 2015
  - Promote adoption of the DICOM standard for Quantitative Imaging applications
  - Develop best practices for storing QI analysis data using DICOM
  - Understand and lower adoption barriers
- Educate vendors so they adopt standards
- Educate customers so they demand standards
- 12 platforms participated or evaluated, including 6 commercial



# Acknowledgments


- Ron Kikinis
- Fiona Fennessy
- Clare Tempany
- Jayashree Kalpathy-Cramer
- Andrew Beers
- Hugo Aerts
- David Clunie
- Steve Pieper
- Reinhard Beichel
- Christian Bauer
- Michael Onken
- Jörg Riesmeier
- Christian Herz
- Ethan Ulrich
- Andreas Wahle
- John Buatti
- John Sunderland

QIICR team

DICOM4QI participants

3D Slicer community

National Cancer Institute

- funding: U24 CA180918, U01 CA151261, U01 CA190234, U24 CA199460, U01 CA190234
- ITCR community
- QIN community