

# Standardized representation of the TCIA LIDC-IDRI annotations using DICOM (DICOM-LIDC-IDRI-Nodules)

## Summary


This dataset contains standardized DICOM representation of the annotations and characterizations collected by the LIDC/IDRI initiative, originally stored in XML and available in the TCIA [LIDC-IDRI](#) collection . Only the nodules that were deemed to be greater or equal to 3 mm in the largest planar dimensions have been annotated and characterized by the expert radiologists performing the annotations. Only those nodules are included in the present dataset.

Conversion was enabled by the *pylidc* library (<https://pylidc.github.io/>) (parsing of XML, volumetric reconstruction of the nodule annotations, clustering of the annotations belonging to the same nodule, calculation of the volume, surface area and largest diameter of the nodules) and the *dcmqi* library (<https://github.com/qiicr/dcmqi>) (storing of the annotations into DICOM Segmentation objects, and storing of the characterizations and measurements into DICOM Structured Reporting objects). The script used for the conversion is available at <https://github.com/qiicr/lidc2dicom>. The details on the process of the conversion and the usage of the resulting objects are available in the preprint citation (see Citations & Data Usage Policy tab).

## Data Access

### Data Access

Click the **Download** button to save a ".tcia" manifest file to your computer, which you must open with the [NBIA Data Retriever](#).

Data Type	Download all or Query/Filter
Structured Reports (SR) and Segmentations (DICOM)	

Please contact [help@cancerimagingarchive.net](mailto:help@cancerimagingarchive.net) with any questions regarding usage.

## Collections Used in this Third Party Analysis

Below is a list of the Collections used in these analyses:

- [LIDC-IDRI](#)

## Detailed Description

### Detailed Description

Image Statistics	
Modalities (DICOM)	Seg, SR
Number of Patients	875
Number of Studies	883
Number of Series	13,718
Number of Images	13,718
Images Size (GB)	2 GB

## Citations & Data Usage Policy

### Citations & Data Usage Policy

Users of this data must abide by the [TCIA Data Usage Policy](#) and the [Creative Commons Attribution 3.0 Unported License](#) under which it has been published. Attribution should include references to the following citations:



#### Data Citation

Fedorov, A., Hancock, M., Clunie, D., Brockhhausen, M., Bona, J., Kirby, J., Freymann, J., Aerts, H.J.W.L., Kikinis, R., Prior, F. (2018). **Standardized representation of the TCIA LIDC-IDRI annotations using DICOM**. The Cancer Imaging Archive. <https://doi.org/10.7937/TCIA.2018.h7umfurq>



#### Publication Citation

Fedorov, A., Hancock, M., Clunie, D., Brochhausen, M., Bona, J., Kirby, J., Freymann, J., Pieper S, Aerts H.J.W.L., Kikinis, R., Prior, F. (2020) **DICOM reencoding of volumetrically annotated Lung Imaging Database Consortium (LIDC) nodules**. Medical Physics Dataset Article. <https://doi.org/10.1002/mp.14445>

### TCIA Citation

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**. Journal of Digital Imaging, 26(6), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>

In addition to the dataset citation above, please be sure to cite the following if you utilize these data in your research:

### Publication Citation

Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, van Beek EJR, Yankelevitz D, et al.: **The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans**. Medical Physics, 38: 915–931, 2011. DOI: <https://doi.org/10.1118/1.3528204>

### Data Citation


Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P. Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Castele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., Clarke, L. P. (2015). **Data From LIDC-IDRI** [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX>

## Other Publications Using This Data

TCIA maintains [a list of publications](#) that leverage TCIA data. If you have a manuscript you'd like to add please [contact the TCIA Helpdesk](#).

### Versions


#### Version 3 (Current): 2020/03/26

Data Type	Download all or Query/Filter
Structured Reports (SR) and Segmentations (DICOM)	

#### What changed:


DICOM objects curated and added to the cancerimagingarchive.net

#### Version 2: 2019/05/14

Data Type	Download all or Query/Filter
Structured Reports (SR) and Segmentations (DICOM)	

What changed: DICOM SEG objects no longer encode empty slices to reduce object size. The coded terms used to describe the nodule annotations now use fewer non-standard (99QIIICR) codes. SegmentLabel attribute is populated in the DICOM SEG objects to list nodule annotation name instead of "Nodule", to help with readability for the user.

#### Version 1: 2018/11/30

Data Type	Download all or Query/Filter
Structured Reports (SR) and Segmentations (DICOM)	

Note: Version 1 of this dataset is currently located in a shared Google Drive folder while undergoing verification. When testing is complete the Google Drive folder will be replaced by a different link to the final dataset. If you identify any issues with the data please report them to the [TCIA Helpdesk](#).