



## **Resources for research on The Cancer Imaging Archive; a catalog of enhanced tools for downloading, visualizing and analyzing TCIA data**

Justin Kirby

Technical Director, Cancer Imaging Informatics Lab

Email: [justin.kirby@nih.gov](mailto:justin.kirby@nih.gov)



**NATIONAL CANCER INSTITUTE**

DCTD Division of Cancer Treatment & Diagnosis

CIP Cancer Imaging Program

**Frederick National Laboratory  
for Cancer Research**

*sponsored by the National Cancer Institute*

**UAMS**

UNIVERSITY OF ARKANSAS  
FOR MEDICAL SCIENCES

# Welcome to The Cancer Imaging Archive

The Cancer Imaging Archive (TCIA) is a service which de-identifies and hosts a large archive of medical images of cancer accessible for public download.

[SUBMIT YOUR DATA](#)

[ACCESS THE DATA](#)

# TCIA components



## Data Collection Centers

- Tools and staffing to support data collection, curation, and de-identification

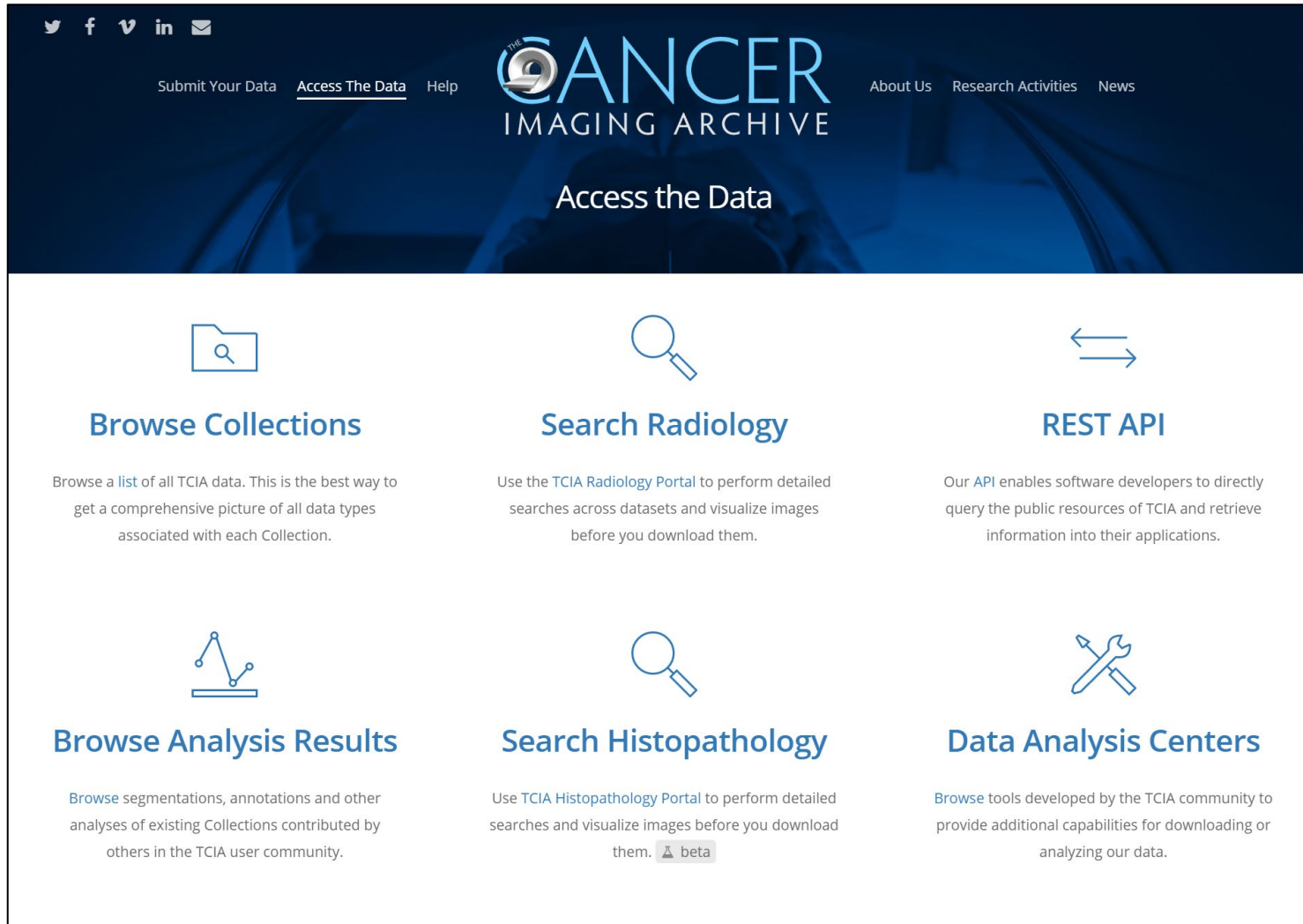
## Data Sets and Access Tools

- Browse
- Filter/Search
- Visualize
- REST API

## Data Analysis Centers

- 3<sup>rd</sup> party web sites or tools which use TCIA's API or mirror its data


# Accessing the Data



[Submit Your Data](#) [Access The Data](#) [Help](#) [About Us](#) [Research Activities](#) [News](#)


**THE CANCER IMAGING ARCHIVE**

## Access the Data




### Browse Collections

Browse a [list](#) of all TCIA data. This is the best way to get a comprehensive picture of all data types associated with each Collection.




### Search Radiology

Use the [TCIA Radiology Portal](#) to perform detailed searches across datasets and visualize images before you download them.




### REST API

Our [API](#) enables software developers to directly query the public resources of TCIA and retrieve information into their applications.




### Browse Analysis Results

Browse segmentations, annotations and other analyses of existing Collections contributed by others in the TCIA user community.



### Search Histopathology

Use [TCIA Histopathology Portal](#) to perform detailed searches and visualize images before you download them. [beta](#)



### Data Analysis Centers

Browse tools developed by the TCIA community to provide additional capabilities for downloading or analyzing our data.

# Browsing the data

TCIA data are organized as "collections"; typically these are patient cohorts related by a common disease (e.g. lung cancer), image modality or type (MRI, CT, digital histopathology, etc) or research focus. Supporting data related to the images such as patient outcomes, treatment details, genomics and image analyses are also provided when available. Try using the filter box above the table to quickly find collections of interest using keywords. Column headers can also be clicked to change the sorting method.

Collection	Cancer Type	Location	Species	Subjects	Image Types	Supporting Data	Access	Status	Updated
Prostate-MRI-US-Biopsy	Prostate Cancer	Prostate	Human	1151	MR, US	Image Analyses	Public	Complete	2020-09-17
QIN-HeadNeck	Head and Neck Carcinomas	Head-Neck	Human	279	PT, CT, SR, SEG, RWV	Clinical, Image Analyses	Public	Complete	2020-09-16
ACRIN-DSC-MR-Brain (ACRIN 6677/RTOG 0625)	Glioblastoma Multiforme	Brain	Human	123	MR, CT	Clinical	Public	Complete	2020-09-09
CRC_FFPE-CODEX_CellNeighs	Colorectal Cancer	Colon	Human	35	Pathology, High-dimensional CODEX images	Clinical, Image Analyses	Public	Complete	2020-09-04
CT Images in COVID-19	COVID-19 (non-cancer)	Lung	Human	632	CT		Public	Complete	2020-08-31
HNSCC-mIF-mHC-comparison	Head and Neck Cancer	Head-Neck	Human	8	Pathology	Image analyses	Public	Complete	2020-07-31

Title	Cancer Type	Location	Subjects	Collections	Analysis Artifacts on TCIA	Updated
DICOM SR of clinical data and measurement for breast cancer collections to TCIA	Breast	Breast	474	TCGA-BRCA, BREAST-DIAGNOSIS, ISPY1, Breast-MRI-NACT-Pilot	DICOM SR descriptions of patient characteristics, histopathology, receptor status and clinical findings including measurements.	2020-05-28
DICOM-SEG Conversions for TCGA-LGG and TCGA-GBM Segmentation Datasets	Glioblastoma, Low Grade Glioma	Brain	167	TCGA-GBM, TCGA-LGG	Tumor segmentations	2020-04-30
Integration of CT-based Qualitative and Radiomic Features with Proteomic Variables in Patients with High-Grade Serous Ovarian Cancer: An Exploratory Analysis	Ovarian	Ovary	20	TCGA-OV	Radiologist assessments of image features, proteogenomic features	2020-04-15
Thoracic Volume and Pleural Effusion Segmentations in Diseased Lungs for Benchmarking Chest CT Processing Pipelines	Lung	Lung	402	NSCLC-Radiomics	Thoracic segmentations, pleural effusion segmentations, image features	2020-04-08
Standardized representation of the TCIA LIDC-IDRI annotations using DICOM	Lung	Chest	1,010	LIDC-IDRI	Tumor segmentations, image features	2020-03-26
Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach	Lung, Head-Neck	Lung, Head-Neck	701	NSCLC-Radiomics, NSCLC-Radiomics-Genomics, Head-Neck-Radiomics-HN1, NSCLC-Radiomics-Interobserver1, RIDER Lung CT	Tumor segmentations and radiomic features	2020-03-23
RIDER Lung CT Segmentation Labels from: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach	Lung	Chest	31	RIDER Lung CT	Tumor segmentations	2020-02-13
Dataset of Segmented Nuclei in Hematoxylin and Eosin Stained Histopathology Images	Various (14 collections)	Various (14 collections)		Various (14 TCGA collections)	Nuclei segmentations	2020-02-08

# Radiology Downloads & Data Portal

## CPTAC-GBM

Created by Tracy Nolan, last modified by natasha.honomichl on Feb 14, 2020

### Summary

This collection contains subjects from the National Cancer Institute's [Clinical Proteomic Tumor Analysis Consortium](#) Glioblastoma Multiforme (CPTAC-GBM) molecular basis of cancer through the application of large-scale proteome and genome analysis, or proteogenomics. Radiology and pathology images from The Cancer Imaging Archive to enable researchers to investigate cancer phenotypes which may correlate to corresponding proteomic, genomic and clinical data.

CPTAC Phase 3 collects data from ten cancer types. In TCIA, imaging from each cancer type will be contained in its own TCIA Collection, with the collection available on TCIA each quarter as it is collected. A summary of CPTAC Phase 3 imaging efforts can be found on the [CPTAC Imaging Proteomics](#) page.

Radiology imaging is collected from standard of care imaging performed on patients immediately before the pathological diagnosis, and from follow-up heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. Pathology imaging is collected as part of the CPTAC qualification effort.

### CPTAC Imaging Special Interest Group

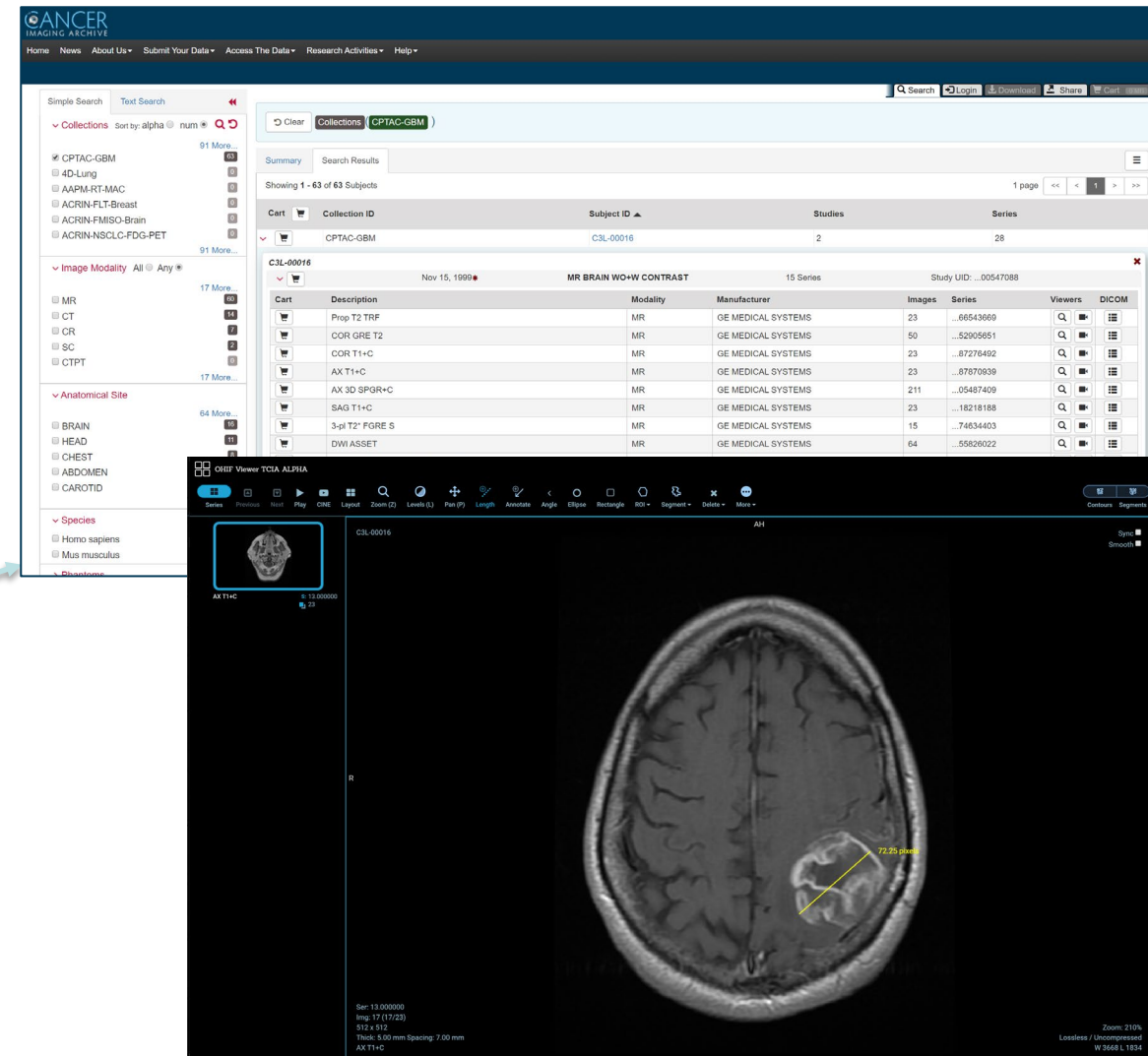
You can join the [CPTAC Imaging Special Interest Group](#) to be notified of webinars & data releases, collaborate on common data wrangling tasks and seek feedback. Webinars such as slide decks and video recordings can be found on the [CPTAC SIG Webinars](#) page.

[Data Access](#) [Detailed Description](#) [Citations & Data Usage Policy](#) [Versions](#)

### Data Access

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

Data Type	Download all or Query/Filter
Images (DICOM, 39.4 GB)	<a href="#">Download</a> <a href="#">Search</a>
Tissue Slide Images (SVS, 87 GB)	<a href="#">Download</a> <a href="#">Search</a>
Clinical Data API (JSON - <a href="#">more info</a> )	<a href="#">Download</a>
Discovery Study Proteomics/Clinical Data (external)	<ul style="list-style-type: none"> <li><a href="#">CPTAC Data Portal (Georgetown)</a></li> <li><a href="#">Proteomic Data Commons</a></li> </ul>
Genomics/Clinical Data (External)	<a href="#">Genomic Data Commons</a>



# Pathology Downloads & Data Portal

## CPTAC-GBM

Created by Tracy Nolan, last modified by natasha.honomichl on Feb 14, 2020

### Summary

This collection contains subjects from the National Cancer Institute's [Clinical Proteomic Tumor Analysis Consortium](#) Glioblastoma Multiforme (CPTAC-GBM) molecular basis of cancer through the application of large-scale proteome and genome analysis, or proteogenomics. Radiology and pathology images from The Cancer Imaging Archive to enable researchers to investigate cancer phenotypes which may correlate to corresponding proteomic, genomic and clinical data.

CPTAC Phase 3 collects data from ten cancer types. In TCIA, imaging from each cancer type will be contained in its own TCIA Collection, with the collection available on TCIA each quarter as it is collected. A summary of CPTAC Phase 3 imaging efforts can be found on the [CPTAC Imaging Proteomics](#) page.

Radiology imaging is collected from standard of care imaging performed on patients immediately before the pathological diagnosis, and from follow-up heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. Pathology imaging is collected as part of the CPTAC qualification

### CPTAC Imaging Special Interest Group

You can join the [CPTAC Imaging Special Interest Group](#) to be notified of webinars & data releases, collaborate on common data wrangling tasks and see webinars such as slide decks and video recordings can be found on the [CPTAC SIG Webinars](#) page.

[Data Access](#) [Detailed Description](#) [Citations & Data Usage Policy](#) [Versions](#)

### Data Access

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

Data Type	Download all or Query/Filter
Images (DICOM, 39.4 GB)	<a href="#">Download</a> <a href="#">Search</a>
Tissue Slide Images (SVS, 87 GB)	<a href="#">Download</a> <a href="#">Search</a>
Clinical Data API (JSON - <a href="#">more info</a> )	<a href="#">Download</a>
Discovery Study Proteomics/Clinical Data (external)	<ul style="list-style-type: none"> <li><a href="#">CPTAC Data Portal (Georgetown)</a></li> <li><a href="#">Proteomic Data Commons</a></li> </ul>
Genomics/Clinical Data (External)	<a href="#">Genomic Data Commons</a>

The screenshot displays the TCIA CPTAC Pathology Portal interface. At the top, it shows 'Selected: 510 slides, Total: 6121 slides' and 'Last Updated: Feb 13, 2020'. The main content is divided into two sections: 'Specimen Data' and 'Slide Browser'.

The 'Specimen Data' table lists the following columns: Case\_ID, Specimen\_ID, Slide\_ID, Tumor, Topographic\_Site, Tumor\_Site, Specimen\_Type, Radiology, Pathology, Genomics, Proteomics, Age, Race, and Ethnicity. The table contains 10 rows of data for various GBM specimens.

The 'Slide Browser' section shows a large histology slide image. The interface includes navigation controls and a search bar. On the left side, there are filters for 'Tumor' (with a bar chart showing counts for various cancer types like AML, CRC, GBM, etc.), 'HasRadiology' (with a bar chart showing 'no' and 'yes' counts), 'Topographic\_Site' (with a list of sites like Bone Marrow, Brain, etc.), and 'Specimen\_Type'.

# Extending the utility of TCIA data

- API enables software developers to access data directly without a web browser
- Data Analysis Centers provide a mechanism to alert others about tools/sites that provide additional ways to access TCIA data
  - Tools that use the API
  - Sites mirroring TCIA datasets

Twitter Facebook YouTube LinkedIn Email

Submit Your Data Access The Data Help

THE **CANCER**  
IMAGING ARCHIVE

About Us Research Activities News

## Access the Data

### Browse Collections

Browse a [list](#) of all TCIA data. This is the best way to get a comprehensive picture of all data types associated with each Collection.

### Search Radiology

Use the [TCIA Radiology Portal](#) to perform detailed searches across datasets and visualize images before you download them.

### REST API

Our [API](#) enables software developers to directly query the public resources of TCIA and retrieve information into their applications.

### Browse Analysis Results

Browse segmentations, annotations and other analyses of existing Collections contributed by others in the TCIA user community.

### Search Histopathology

Use [TCIA Histopathology Portal](#) to perform detailed searches and visualize images before you download them. [beta](#)

### Data Analysis Centers

Browse tools developed by the TCIA community to provide additional capabilities for downloading or analyzing our data.



# Data Analysis Center Categories

- Tools you can install yourself
  - Desktop applications
  - Web applications
- Code sharing
  - Github repositories
  - Coding tutorials/notebooks
- 3<sup>rd</sup> Party Sites/Services
  - Groups hosting copies of our datasets their own websites

The screenshot shows the TCIA website interface. At the top, there are social media icons (Twitter, Facebook, YouTube, LinkedIn, Email) and navigation links: "Submit Your Data", "Access The Data", "Help", "About Us", "Research Activities", and "News". The main header features the TCIA logo and the text "Access the Data". Below this, there are six main categories, each with an icon and a brief description:

- Browse Collections**: Browse a list of all TCIA data. This is the best way to get a comprehensive picture of all data types associated with each Collection.
- Search Radiology**: Use the TCIA Radiology Portal to perform detailed searches across datasets and visualize images before you download them.
- Browse Analysis Results**: Browse segmentations, annotations and other analyses of existing Collections contributed by others in the TCIA user community.
- Search Histopathology**: Use TCIA Histopathology Portal to perform detailed searches and visualize images before you download them. beta
- REST API**: Our API enables software developers to directly query the public resources of TCIA and retrieve information into their applications.
- Data Analysis Centers**: Browse tools developed by the TCIA community to provide additional capabilities for downloading or analyzing our data.

The "REST API" and "Data Analysis Centers" sections are highlighted with a red border in the original image.

# Data Analysis Center Catalog

## Data Analysis Centers (DACs)

Created by Justin Kirby, last modified on Oct 15, 2021

A Data Analysis Center (DAC) is a tool or website which provides additional capabilities for downloading, visualizing or analyzing TCIA data by connecting to our [TCIA Programmatic Interface \(REST API\)](#) or by mirroring our [Collections](#). If you have developed something which meets these criteria please [contact the helpdesk](#) so we can add it to this page. We will also work with you to ensure your site/tool provides adequate attributions and links back to TCIA to comply with our [Data Usage Policies and Restrictions](#).

Resource	Description	Functionality	TCIA Data Access	Platform
<a href="#">3D Slicer TCIA Browser extension</a>	3D Slicer ( <a href="http://slicer.org">http://slicer.org</a> ) is a free and open source platform for medical image visualization and quantitative analysis. The <a href="#">TCIA Browser extension</a> of 3D Slicer enables integration of the versatile visualization and computing tools of 3D Slicer with unique data resources of TCIA. Among other capabilities, 3D Slicer enables 2-, 3-, and 4-d visualization tools, DICOM interoperability for both images and image annotations, radiomics feature calculation, multi-modality fusion and deformable registration, a collection of segmentation tools, Matlab and python interface, and integration of such libraries as ITK, VTK, DCMTK and numpy.	Visualization and Analysis	API	Windows, Mac OS X, Linux
<a href="#">CancerImagingArchive.jl</a>	Julia interface for exploring and downloading data on The Cancer Imaging Archive (TCIA)	Data access	API	Windows, Mac OS X, Linux
<a href="#">Clara Train for TCIA Datasets</a>	<a href="#">Clara Train</a> is NVIDIA's domain-optimized application-development framework for medical-imaging researchers and artificial intelligence (AI) developers. Clara Train SDK, which you deploy in a highly available (HA) configuration on the AWS Cloud, includes an AI Assisted Annotation developer toolkit that can be integrated into existing medical viewers, accelerating the creation of AI-ready, annotated medical-imaging datasets. Clara Train also provides a TensorFlow-based training framework with domain-specific pretrained models that accelerate AI development with techniques like transfer learning, federated learning, and automated machine	Data access, Visualization, and Analysis	API	Linux

# DAC Catalog – Over 20 projects & Code Repositories

## Data Analysis Centers (DACs)

Updated with King, December modification March 10, 2021

A Data Analysis Center (DAC) is a tool or website which provides additional capabilities for downloading, visualizing or analyzing TCGA data by connecting to our [TCGA Programmatic Interface \(REST API\)](#) or by mirroring our [Subsites](#). If you have developed something which meets these criteria please [contact us](#) so we can add it to this page. We will also work with you to ensure your site/tool provides adequate attributions and links back to TCGA to comply with our [Data Usage Policies and Restrictions](#).

Resource	Description	Functionality	TCGA Data Access	Platform
<a href="#">3D Slicer TCGA Browser Extension</a>	3D Slicer ( <a href="http://slicer.org">http://slicer.org</a> ) is a free and open source platform for medical image visualization and quantitative analysis. The <a href="#">TCGA Browser extension</a> of 3D Slicer enables integration of the versatile visualization and computing tools of 3D Slicer with unique data resources of TCGA. Among other capabilities, 3D Slicer enables 2-, 3-, and 4-d visualization tools, DICOM interoperability for both images and image annotations, radiomics feature calculation, multi-modality fusion and deformable registration, a collection of segmentation tools, Matlab and python interface and integration of such libraries as ITK, VTK, DCMIO and numpy.	Visualization and Analysis	API	Windows, Mac OS X, Linux
<a href="#">CancerImagingArchives.j</a>	Julia interface for exploring and downloading data on The Cancer Imaging Archive (TCIA)	Data access	API	Windows, Mac OS X, Linux
<a href="#">Data Train for TCIA Datasets</a>	Data Train is NVIDIA's domain-optimized application-development framework for medical-imaging researchers and artificial intelligence (AI) developers. Data Train SDK, which you deploy in a highly available (HA) configuration on the AWS Cloud, includes an AI-Assisted Annotation developer toolkit that can be integrated into existing medical viewers, accelerating the creation of AI-ready, annotated medical-imaging datasets. Data Train also provides a TensorFlow-based training framework with domain-specific pretrained models that accelerate AI development with techniques like transfer learning, federated learning, and automated machine learning. Models trained with Data Train are packaged as Medical Model Archives (MMARs), which provide a standardized format for training workflows and collaborations. A sample notebook has been created in order to easily import and pre-process TCIA data for analysis with Data Train.	Data access, Visualization, and Analysis	API	Linux
<a href="#">Community Code Share on GitHub</a>	If you've developed open source code you'd like to share with the community you can use GitHub's <a href="#">topic feature</a> to make it discoverable by tagging it with "tcga-dac". Please note these tools are not directly supported by TCGA or its partners.	Data access, Visualization, and Analysis	API / Mirrored	Miscellaneous
<a href="#">DataScope</a>	An open source data exploration and visual analytic tool that uses a declarative grammar to author interactive dashboards. Using a series of JSON files that describe the data, we are able to fuse clinical, radiology and digital pathology data. The <a href="#">TCGA CFAC Pathology Portal</a> is powered by DataScope.	Data access, Visualization	API	Web application
<a href="#">eRAD</a>	eRAD is a freely available quantitative imaging informatics platform, developed by the Rubin Lab at <a href="#">Stanford Medicine Radiology</a> , at <a href="#">Stanford University</a> . Its built-in connection to our REST API allows TCIA data to be seamlessly imported into eRAD for analysis.	Visualization and Analysis	API	Web application
<a href="#">i2-DOC Plus</a>	The Georgetown Database of Cancer Plus other diseases (i2-DOC Plus) is a precision medicine platform containing molecular and clinical data from thousands of patients and cell lines, along with tools for analysis and data visualization. It contains mirrored data from the <a href="#">IMMORT-DIAGNOSIS</a> collection.	Visualization and Analysis	Mirrored	Web application
<a href="#">Google Cloud Healthcare API</a>	The Cloud Healthcare API provides access to TCIA datasets via Google Cloud Platform (GCP) from Cloud Storage, BigQuery, or using the Cloud Healthcare API, as described in <a href="#">GCP data access</a> .	Data Access	Mirrored	Web application
<a href="#">MONAI</a>	MONAI ( <a href="https://monai.io">https://monai.io</a> ) is a freely available, community-supported, PyTorch-based framework for deep learning in healthcare imaging. It provides deep learning capabilities tailored for healthcare imaging research, development and deployment. MONAI is part of the larger Project MONAI effort, that includes <a href="#">MONAI Label</a> for AI-assisted annotation, <a href="#">MONAI Deploy</a> for packaging, distributing, and deploying MONAI-based applications, and <a href="#">MONAI Tutorials</a> which contains extensive educational and community building resources. MONAI's <a href="#">TCIA Connect Tutorial</a> describes how patient data (images, seg results, etc.) from NCJ repositories such as the <a href="#">DOC</a> and <a href="#">TCIA</a> can be leveraged for MONAI model development and deployment. MONAI is available on cloud services such as Google Cloud and Amazon SageMaker and access every major operating system and Python version via pip. MONAI Label is also available as a 3D Slicer Plugin.	Data access, Visualization, and Analysis	API	Windows, Mac OS X, Linux
<a href="#">NCI Imaging Data Commons</a>	NCI Imaging Data Commons (IDC) is a cloud-based resource within NCI Cancer Research Data Commons (CRDC) that connects researchers with cancer imaging datasets, resources for exploring those datasets and identifying relevant cohorts, and other components of CRDC that will host additional data types and support computation on the defined cohorts.	Visualization and Analysis	Mirrored	Cloud-based platform
<a href="#">OncoRx Medical TCIA BioRxiv app</a>	Repository with minimal docker compose configuration and script to create a DICOM server with a TCIA collection locally. Can be extended modularly with additional docker images for deep learning experiments.	Data Access	API	Windows, Mac OS X, Linux
<a href="#">Orthanc TCIA Plugin</a>	This plugin extends Orthanc with a Web interface that can be used to import open-data medical images from The Cancer Imaging Archive (TCIA), and serve them immediately using Orthanc.  The plugin can be used to import so-called "cart spreadsheets" generated by the <a href="#">NCIA Search Client</a> , or to browse the image collections of TCIA marks to its REST API.	Data Access	API / Mirrored	Windows, Mac OS X, Linux
<a href="#">REDI PathDB</a>	REDI Pathology Data Management Prototype for TCIA	Data access, Visualization, and Analysis	API	Web application
<a href="#">researchcanai</a>	Research-MedicalImaging (Research-RI) is an open-source, web-based platform which enables deployment of AI models while simultaneously providing standard image viewing and reporting schemes. The goal of Research-RI is to ingest 3D medical imaging and provide a 4th dimension (AI) when requested by a user. As a case study, we demonstrate the utility of our platform and present <a href="#">ProstateCancer.ai</a> (see also <a href="https://github.com/researchcanai/prostatecancer.ai">https://github.com/researchcanai/prostatecancer.ai</a> ), a web application which uses data from <a href="#">SPR-KAPPA-NCI-PAD(DAT)S Challenges (PROSTATIS)</a> for identification of clinically significant prostate cancer in MRI. The user can put the AI-assisted probe at any location on the images to see the result of the AI prediction for that specific location. For the reporting, the user can utilize the <a href="#">Prostate-4D</a> interface which is provided. All the user's annotations will be saved in a database for further analysis.	Visualization and Analysis	Mirrored	Web application
<a href="#">slicer</a>	<a href="#">slicer</a> is a python library intended to improve workflow associated with the <a href="#">i2-DOC</a> dataset.	Visualization and Analysis	N/A	Windows, Mac OS X, Linux
<a href="#">Seven Bridges Cancer Genomics Cloud (CGC)</a>	An NCI-funded platform that is available to any non-commercial researcher for cloud-based data access and analysis. Through the CGC, users can access petabytes of public data, including select collections from TCIA, as well as hundreds of bioinformatic tools and workflows for scalable, cost-effective analysis in the cloud alongside their own data.	Data Access, Visualization, Analysis	Mirrored	Web application
<a href="#">tortoiseshell</a>	This Python package uses the official TCIA REST API to enable downloads from <a href="http://www.cancerimagingarchive.net">www.cancerimagingarchive.net</a> from within Python scripts and Jupyter notebooks. The documentation can be found at <a href="https://tortoiseshell.pygithub.io/tortoiseshell/">https://tortoiseshell.pygithub.io/tortoiseshell/</a> . This PyPi/Conda package is based on source code of the TCIA API SDK <a href="https://github.com/TCIA-Community/TCIA-API-SDK">https://github.com/TCIA-Community/TCIA-API-SDK</a> .	Data Access	API	Windows, Mac OS X, Linux
<a href="#">TCIApathfinder</a>	A user-friendly R client for the TCIA REST API	Data access	API	Windows, Mac OS X, Linux
<a href="#">TCIA Python3 Downloader</a>	A python3 client designed to provide users of The Cancer Imaging Archive with the ability to easily interact and download data from the TCIA Programmatic Interface (REST API).	Data Access	API	Windows, Mac OS X, Linux
<a href="#">Zigami</a>	Zigami helps easily find patterns, outliers and trends in large, curated image data sets, and uncover bias, overfitting and misclassifications in machine learning models, to assist with providing explainability of your Machine Learning models.  Our scalable, cloud-based platform is powered by an image rendering engine and based on gaming technology. It can display tens of thousands of images (static or dynamic) over low bandwidth connections, and supports a wide variety of image and video formats.  Our solution helps with:	Visualization, Analysis	Mirrored	Web application

# ORTHANC - PACS Integration

## ORTHANC Official Plugin

File Edit View History Bookmarks Tools Help

TCIA for Orthanc

localhost:8042/tcia/app/index.html

**THE CANCER IMAGING ARCHIVE**  
**ORTHANC**

### Import NBIA cart

Import into Orthanc the content of a cart that has been exported as a spreadsheet from the [NBIA Search Client](#) (cf. [screenshot](#)).

Choose file Browse

Submit

### Explore TCIA collections

Note that loading the information about the TCIA collections takes time because of the throttling that is applied to the [REST API of TCIA](#). Orthanc caches this information in memory to speed up further access to the same resources. This cache can be cleared using the button at the bottom of this page.

Filter:

Collection	Modalities	Body parts	
TCGA-GBM	CT, DX, MR, SEG	BRAIN	<span>▶</span>
LIDC-IDRI	CR, CT, DX, SEG, SR	CHEST	<span>▶</span>
BREAST-DIAGNOSIS	CT, MG, MR, PT, SR	BREAST	<span>▶</span>
PROSTATE-MRI	MR	PROSTATE	<span>▶</span>
QIN-BREAST	CT, MR, PT	BREAST	<span>▶</span>

## Oncora Medical “TCIA Bootstrap” w/ ORTHANC

github.com/oncoramedical/tcia\_bootstrap

README.md

## TCIA Bootstrap

### Introduction

A small docker set up to pull TCIA collections into a DICOM server.

### Pre-requisites

Make sure that the following are installed:

1. [Docker](#)
2. [Docker compose](#)

### Directions

### TL;DR

```
docker-compose up
bash bootstrap.sh -h
bash bootstrap.sh -c <COLLECTION>
```

# 3D Slicer - Image Analysis Integration

TCIA Browser

Current Collection: TCGA-GBM  Cache server responses

Patients (Accessed: Sun May 17 23:47:28 2015)

Patient ID	Patient Name	Patient BirthDate	Patient Sex	Ethnic Group	Clinical Data
1	TCGA-06-0119	TCGA-06-0119	F		
2	TCGA-06-0122	TCGA-06-0122	F		
3	TCGA-06-0125	TCGA-06-0125	F		
4	TCGA-06-0127	TCGA-06-0127	M		
5	TCGA-06-0128	TCGA-06-0128	M		

Studies

Study Date	Study Description	Admitting Diagnosis Description	Study ID	Patient Age	Series Count
1 2003-12-26					16
2 2000-03-07					11
3 2001-09-11					11
4 2002-06-18					14
5 2000-03-07					11

Select: All None

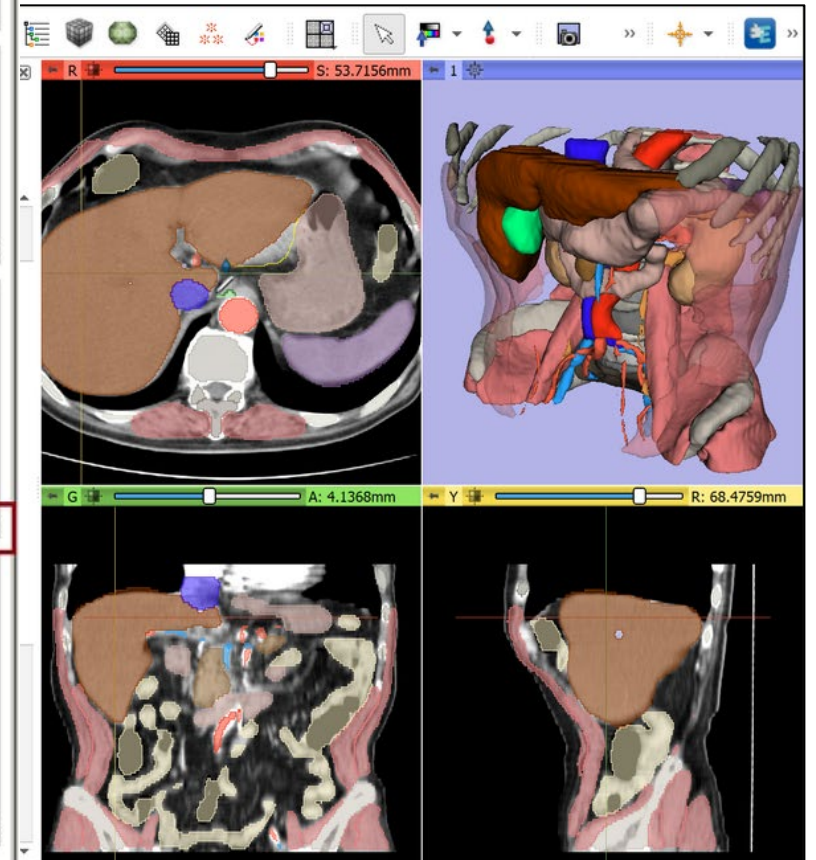
Series (Accessed: Sun May 17 23:47:41 2015)

Status	Modality	Protocol Name	Series Date	Series Description	Body Part Examined	Series Number	Ant
1	MR		2003-12-26	SAG T1	BRAIN	2.000000	
2	MR		2003-12-27	+COR T1	BRAIN	12.000000	
3	MR		2003-12-27	SAG T1	BRAIN	4.000000	
4	MR		2003-12-27	AXIAL T1	BRAIN	6.000000	

Select: All None

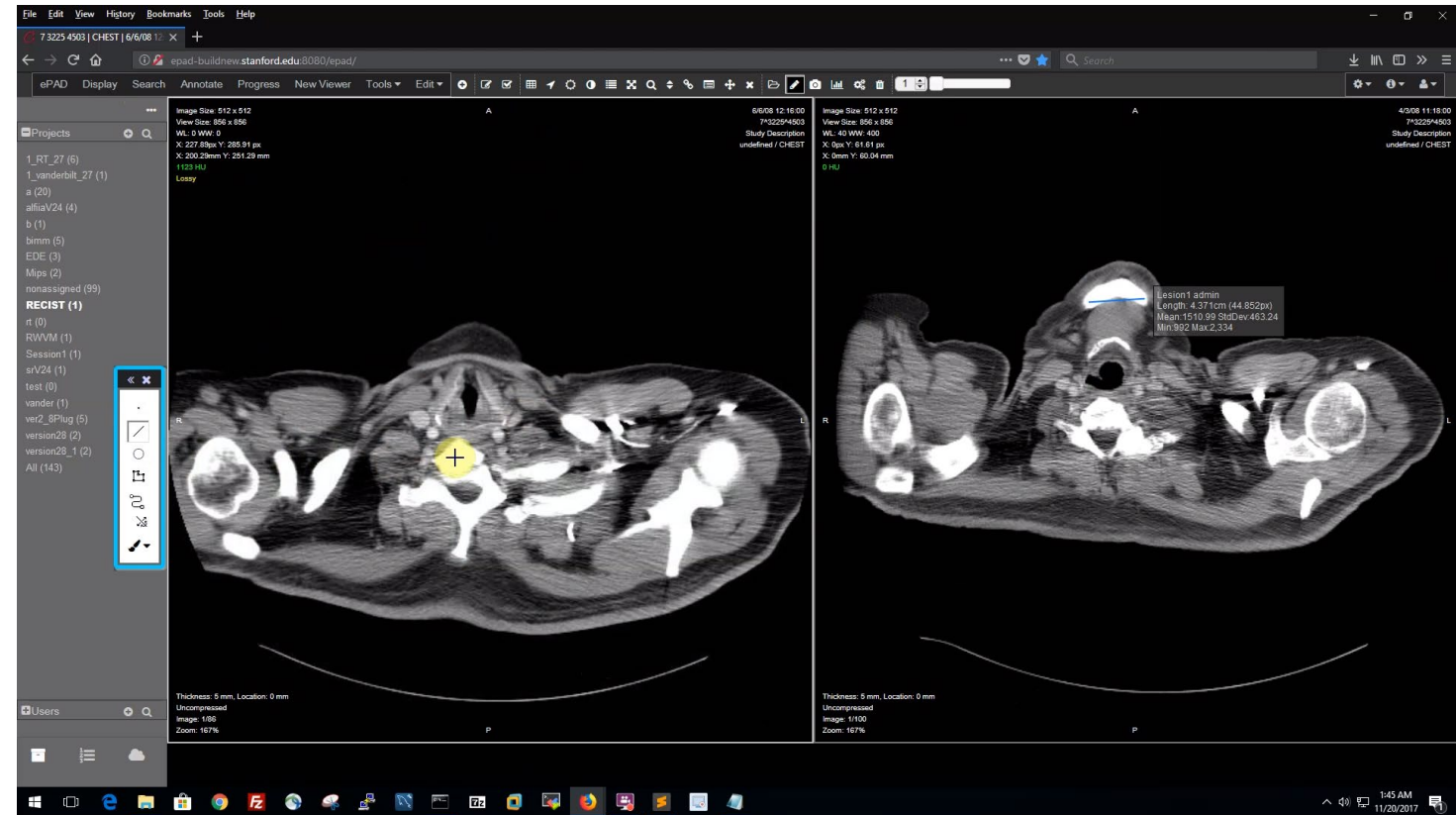
No. of images to download: 46

F G



# Epad - PACS and Image Analysis Integration

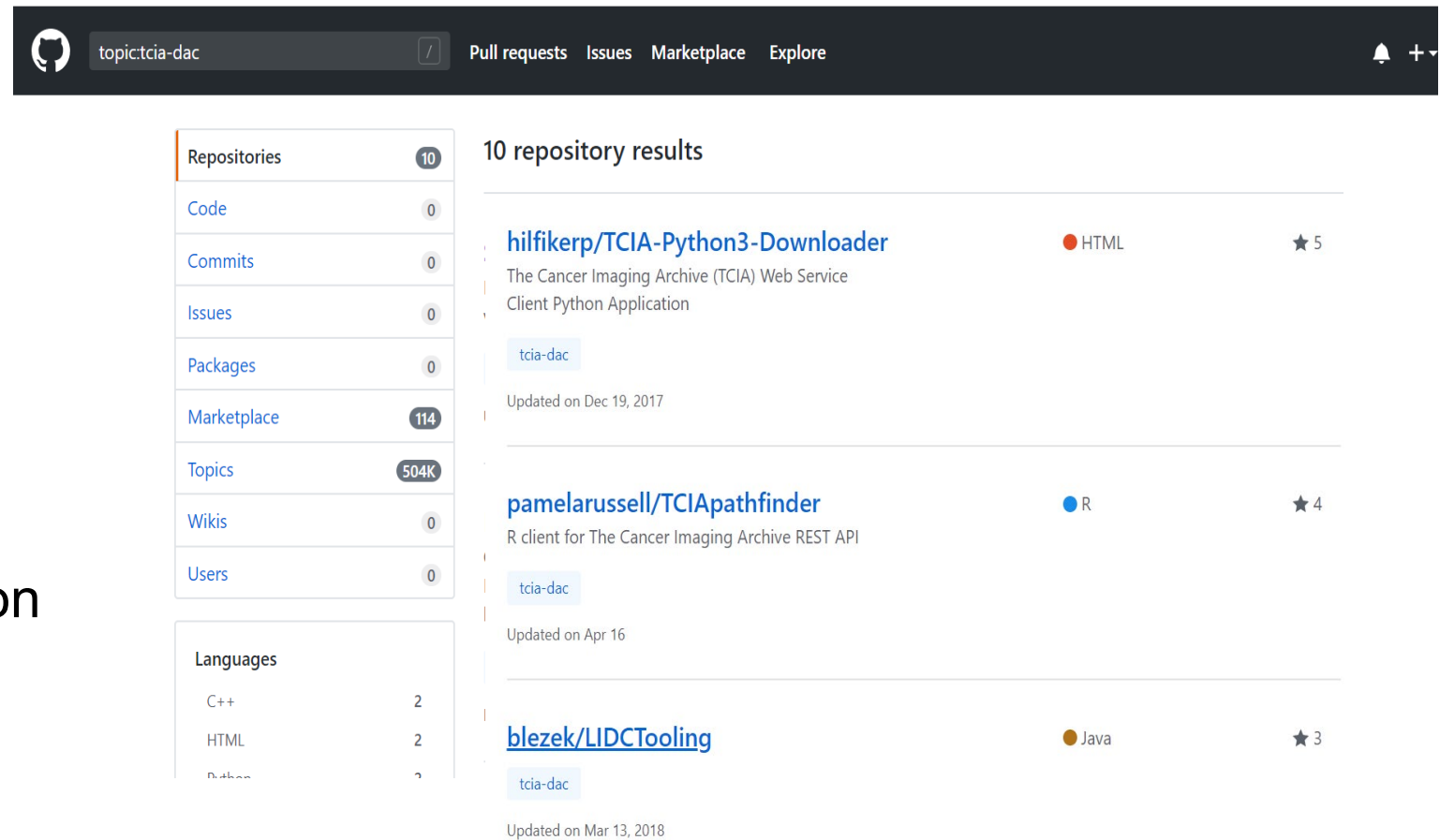
- Uses API to easily import images from TCIA into a PACS back end
- Allows visualization, annotation/segmentation, and structured data capture with case report forms and controlled ontologies



# TCIA community code sharing

- API wrappers for popular programming languages
  - Python
  - R
  - Julia
- Specialized code for working with specific datasets
  - Visualizing custom annotation formats

Tag your Github repo with “tcia-dac” topic tag to appear in the list



The screenshot shows a GitHub search results page for the topic "tcia-dac". The search bar at the top contains "topic:tcia-dac". The page displays 10 repository results. On the left, there are navigation menus for Repositories (10), Code (0), Commits (0), Issues (0), Packages (0), Marketplace (114), Topics (504K), Wikis (0), and Users (0). Below these are Language filters for C++ (2), HTML (2), and Python (2). The main content area shows three repository results:

Repository Name	Language	Stars
<a href="#">hilfikerp/TCIA-Python3-Downloader</a>	HTML	5
<a href="#">pamelarussell/TCIApathfinder</a>	R	4
<a href="#">blezek/LIDCTooling</a>	Java	3

Each result includes a "tcia-dac" topic tag and an "Updated on" date.

# Code Sharing – Reproducible Research

☰ README.rst

## Brain Tumor Progression Dataset

### Files

<b>Directories:</b>	Directories 01 - 11 include pairs of Pre and Post T1 MRI Scans and Masks (four per directory)
<b>supplementary.csv:</b>	CSV file that gives additional information for the data and includes details of what slices were removed in order to transform the original data into this dataset. This file is used by <code>process.py</code> (explained below) to modify the original dataset into this dataset.
<b>process.py:</b>	Python script to modify original dataset into this dataset, contains useful helper functions to use in the interpreter or script to make further modifications. This file is messy and unneeded but gives a technical description of the changes made to supplement the explanation below.
<b>metadata.csv:</b>	The metadata file from the TCIA downloader showing the raw data this dataset is based on/downloaded from.
<b>licence.html:</b>	The licence file from the TCIA downloader. This is the original licence file that is included when the data is downloaded from the source.
<b>README.rst:</b>	This document

### Data References and Usage

This repository includes data from the Brain-Tumor-Progression collection published by The Cancer Imaging Archive. The original data has been modified in the following ways: Firstly all DICM image slices per set have been combined into one Gzipped NifTI file, renaming the resulting NifTI images Pre.... and Post... for the first and second scan respectively, for both the MRI Scans and associated tumor masks resulting in four .nii.gz images. The images have all been resized to 256 X 256 using Nearest Neighbors interpolation and the number of slices was first reduces to 22, for some sets that required a few slides to be dropped and resulted in a closer relationship between Pre and Post slices for the same slice number, for others no change was required. Following that the slices were then padded with blank slides to bring the final count to 24 slices each; this was done so that three subsample convolutional layers could be composed without ending up with decimal dimensions ( $24 \% 2^3 = 0$ ).

The screenshot shows a GitHub repository page for 'SagaraBattousai / BrainTumorProgression'. The repository is public and has 1 branch and 0 tags. The main content area displays a file tree with 11 directories (01-11) and several files (.gitignore, LICENSE, README.rst, licence.html, metadata.csv, process.py, and supplementary.csv). The files are all dated '5 months ago'. On the right side, there is an 'About' section with a description 'No description or website provided.', a 'Readme' link, and a 'View license' link. Below that, there are sections for 'Releases' (No releases published), 'Packages' (No packages published), and 'Languages' (Python 100.0%).



# Jupyter Notebooks - Coding Tutorials

## NVIDIA Clara SDK Tutorial

github.com/NVIDIA/clara-train-examples/blob/master/PyTorch/NoteBooks/Data/TCIA/TCIADownloader.ipynb

### Download dataset form NIH cancer imaging archive

This cancer imaging archive is a great source for medical data sets. This notebook will walk you through: 1- Selecting a dataset 2- Downloading data dicom images to nifti 4- Converting data dicom seg images to nifti 5- Creating your dataset json to begin training your AI models

### 0. Prerequisite

#### 0.1 Clone code

Clone code from tcia\_downloader repo, move it to this directory

```
In [ ]: git clone https://github.com/lescientifik/tcia_downloader /claraDevDay/Data/TCIA/tmp
        mv -vn /claraDevDay/Data/TCIA/tmp/* /claraDevDay/Data/TCIA/
        mv -vn /claraDevDay/Data/TCIA/tmp/src/* /claraDevDay/Data/TCIA/src/
```

```
In [ ]: DataRoot="/claraDevDay/Data/"
        CodeRoot="/claraDevDay/Data/TCIA"
        %cd $CodeRoot
        pwd
```

#### 0.2 Download needed conversion tool

```
In [ ]: wget https://github.com/QIICR/dcmqi/releases/download/v1.2.2/dcmqi-1.2.2-linux.tar.gz && \
        tar xf dcmqi-1.2.2-linux.tar.gz && \
        cp dcmqi-1.2.2-linux/bin/segimage2itkimage /usr/local/bin/ && \
        rm -rf dcmqi-1.2.2-linux*
```

```
In [ ]: curl -fLO https://github.com/rordenlab/dcm2nix/releases/latest/download/dcm2nix_lnx.zip
        unzip dcm2nix_lnx.zip -d /usr/local/bin/
        chmod +x /usr/local/bin/dcm2nix
```

### 1. Select images to download

You can find out different studies from [their site](#). Using the [online tool](#), you can download list of images you would like to download.



Home News About Us Submit Your Data Access The Data Research Activities Help

github.com/Project-MONAI/tutorials/blob/82e1e623c2cafd3b3dd94db537bb743dce523a6/modules/tcia\_csv\_processing.ipynb

### Download raw DICOM series at runtime and execute transform

In this demo, we extract the series UID from loaded TCIA table and try to download the DICOM raw images with REST API, more details: <https://wiki.cancerimagingarchive.net/display/Public/TCIA+REST+API+Guide>

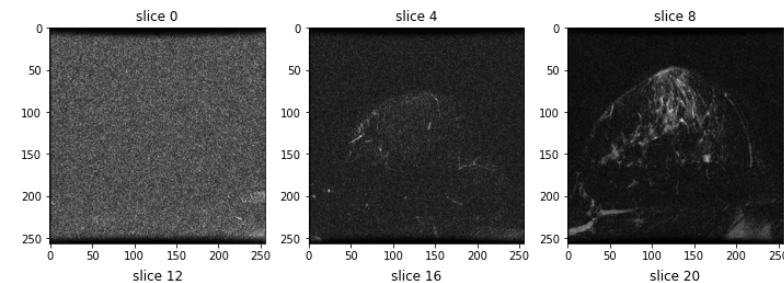
```
In [6]: data = dataset[0]
        print(f"Series UID is: {data['Series UID']}")
        1.3.6.1.4.1.14519.5.2.1.7695.1700.506244151085816331158034893215.zip: 1.74MB [00:04, 379kB/s]
        Downloaded: /workspace/data/medical/tcia_images/1.3.6.1.4.1.14519.5.2.1.7695.1700.506244151085816331158034893215.zip
        Expected md5 is None, skip md5 check for file /workspace/data/medical/tcia_images/1.3.6.1.4.1.14519.5.2.1.7695.1700.506244151085816331158034893215.zip.
        Writing into directory: /workspace/data/medical/tcia_images/1.3.6.1.4.1.14519.5.2.1.7695.1700.506244151085816331158034893215.
        Series UID is: 1.3.6.1.4.1.14519.5.2.1.7695.1700.506244151085816331158034893215
```

### Plot the loaded DICOM series content

Here we access the loaded DICOM image by key image, then plot several slices.

```
In [7]: img = data["image"]
        print(f"image shape: {img.shape}, min value: {img.min()}, max value: {img.max()}")
        plt.figure("check", (12, 8))
        for i in range(6):
            plt.subplot(2, 3, i + 1)
            plt.title(f"slice {i * 4}")
            plt.imshow(img[:, :, i * 4], cmap="gray")
        plt.show()
```

image shape: (256, 256, 24), min value: 0.0, max value: 1624.0



MONAI Tutorial

# Datascope – Multi-modal Data Connectivity

The image displays a multi-modal data connectivity interface. At the center is a data table with the following columns: Case ID, Project, Primary Site, Gender, Age, Race, and various data categories (Radiology, Pathology, Genomics, etc.). A red circle highlights a row with Case ID C3L-00019 and Project CPTAC-3.

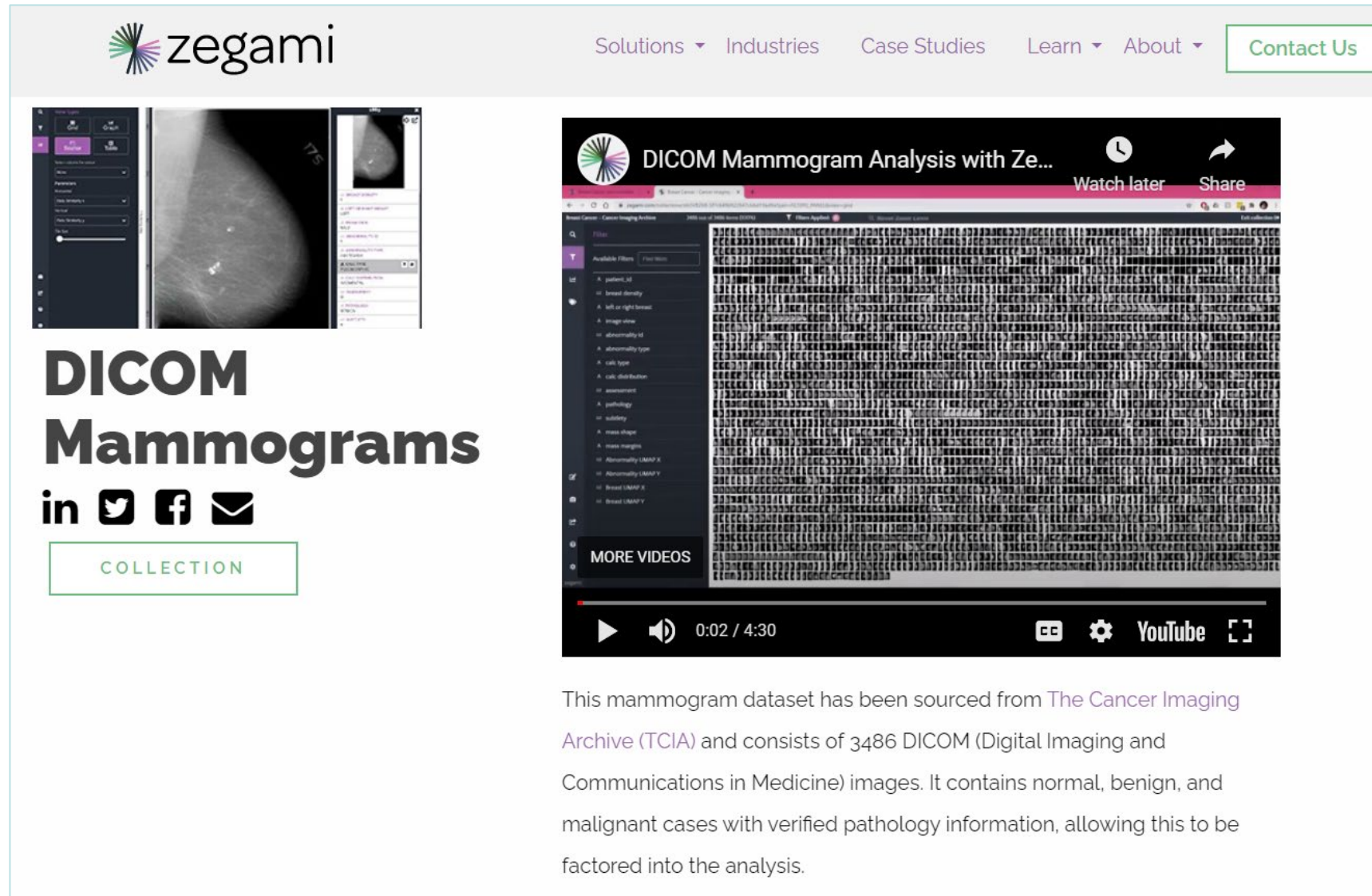
Case ID	Project	Primary Site	Gender	Age	Race	Radiology	Pathology	Genomics
C3L-00016	CPTAC-3	Brain	Male	40-50	White	C3L-00016	C3L-00016-23	
C3L-00016	CPTAC-3	Brain	Male	40-50	White	C3L-00016	C3L-00016-21	
C3L-00016	CPTAC-3	Brain	Male	40-50	White	C3L-00016	C3L-00016-24	
C3L-00019	CPTAC-3	Brain	Female	40-50	White	C3L-00019	C3L-00019-21	
C3L-00019	CPTAC-3	Brain	Female	40-50	White	C3L-00019	C3L-00019-22	
C3L-00019	CPTAC-3	Brain	Female	40-50	White	C3L-00019	C3L-00019-23	
C3L-00019	CPTAC-3	Brain	Female	40-50	White	C3L-00019	C3L-00104-21	C3L-00104
C3L-00019	CPTAC-3	Brain	Female	40-50	White	C3L-00019	C3L-00104	C3L-00104
C3L-00265	CPTAC-3	Brain	Male	60-70	White	C3L-00265	C3L-00265-24	
C3L-00265	CPTAC-3	Brain	Male	60-70	White	C3L-00265	C3L-00265-21	
C3L-00265	CPTAC-3	Brain	Male	60-70	White	C3L-00265	C3L-00265-22	
C3L-00265	CPTAC-3	Brain	Male	60-70	White	C3L-00265	C3L-00265-23	
C3L-00278	CPTAC-3	Brain	Male	60-70	White	C3L-00278	C3L-00278-23	
C3L-00278	CPTAC-3	Brain	Male	60-70	White	C3L-00278	C3L-00278-21	
C3L-00278	CPTAC-3	Brain	Male	60-70	White	C3L-00278	C3L-00278-22	
C3L-00349	CPTAC-3	Brain	Female	50-60	White	C3L-00349	C3L-00349-23	

Surrounding the table are several other data visualization components:

- TCIA CPTAC Tumor**: A sidebar showing a list of tumor types including AML (72), CCRCO (CM) (400), GBM (510), HNSCC (14), and LUSC (14).
- NCI GDC Data Portal**: A search interface with filters for Case, Primary Site (kidney, brain, uterus\_nos), Program (CPTAC), and Disease Type (gliomas, not applicable).
- Proteomic Data Commons**: A dashboard with filters for Project (CPTAC2, CPTAC3), Primary Site (Brain, Breast, Colon, Kidney, Liver, Ovary, Rectum, Uterus\_NOS), and Program (CPTAC-3).
- Topographic\_Site**: A list of anatomical sites including Bone Marrow (0), Brain (510), Head, Neck (0), Kidney (0), Lung (0), Pancreas (0), Skin (0), Soft Tissue (0), and Uterus Endometrium (0).
- Imaging**: A histology slide and an MRI brain scan showing a tumor in the brain.




# Zegami

- Quickly visualize the data based on image characteristics (e.g. similarity) or supporting data (e.g. diagnosis) to discover patterns and outliers
- Watch their demo with the TCIA CBIS-DDSM mammogram dataset at <https://youtu.be/xKJRIHbm50>



The image shows a composite of two screenshots. The top screenshot is the Zegami website's navigation bar, featuring the Zegami logo (a stylized starburst) and the name 'zegami'. To the right are navigation links: 'Solutions', 'Industries', 'Case Studies', 'Learn', and 'About', followed by a green 'Contact Us' button. Below the navigation bar is a screenshot of a software interface displaying a mammogram image with various analysis tools and filters. The bottom screenshot is a YouTube video player titled 'DICOM Mammogram Analysis with Ze...'. The video player shows a list of filters on the left side, including 'patient\_id', 'breast density', 'left or right breast', 'image view', 'abnormality id', 'abnormality type', 'calc type', 'calc distribution', 'assessment', 'pathology', 'subtype', 'mass shape', 'mass margin', 'Abnormality UMMP X', 'Breast UMMP X', and 'Breast UMMP Y'. The video player also shows a 'COLLECTION' button and a 'MORE VIDEOS' button. The video player controls at the bottom indicate the video is at 0:02 / 4:30.

**DICOM Mammograms**

in   

COLLECTION

MORE VIDEOS

0:02 / 4:30

YouTube

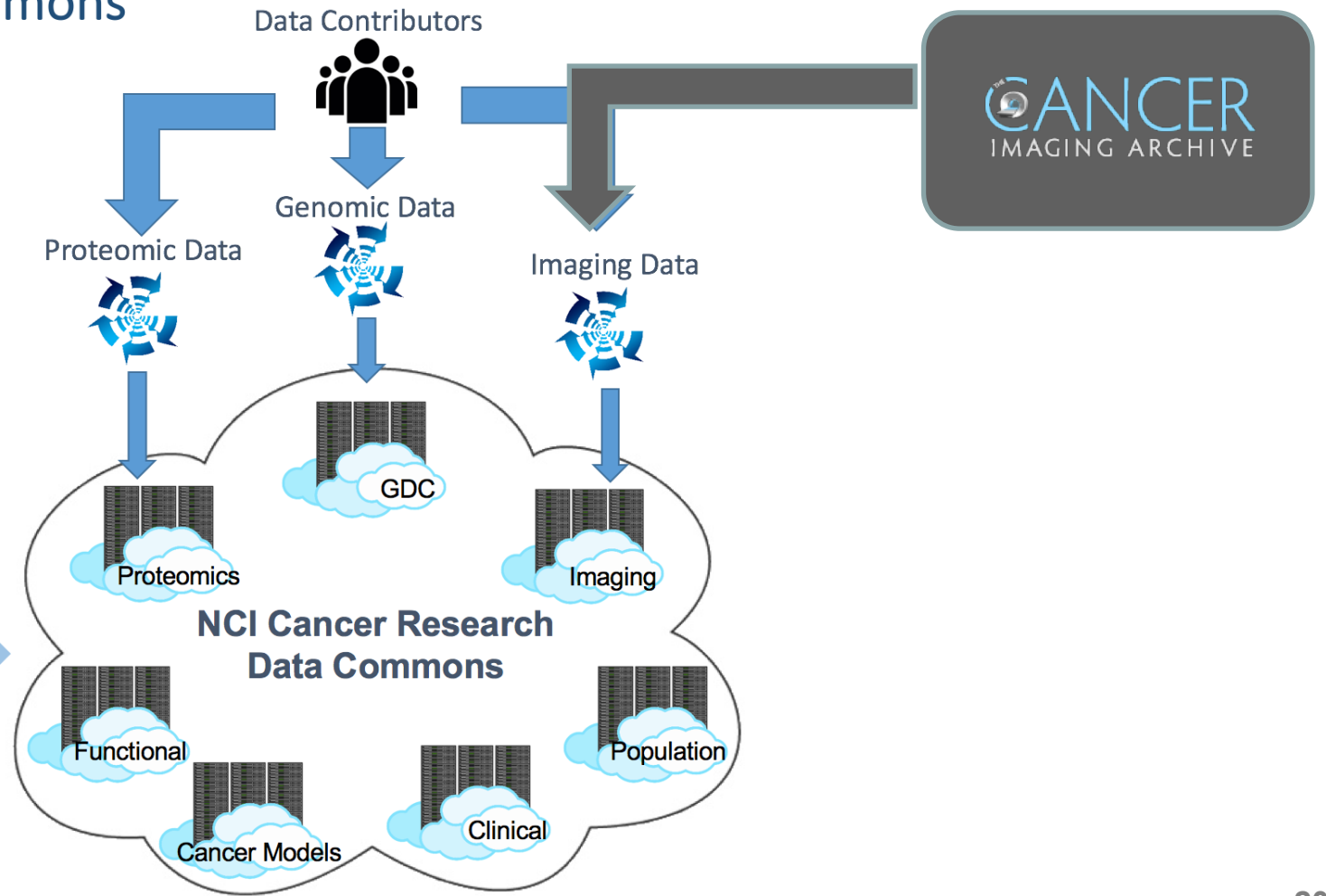
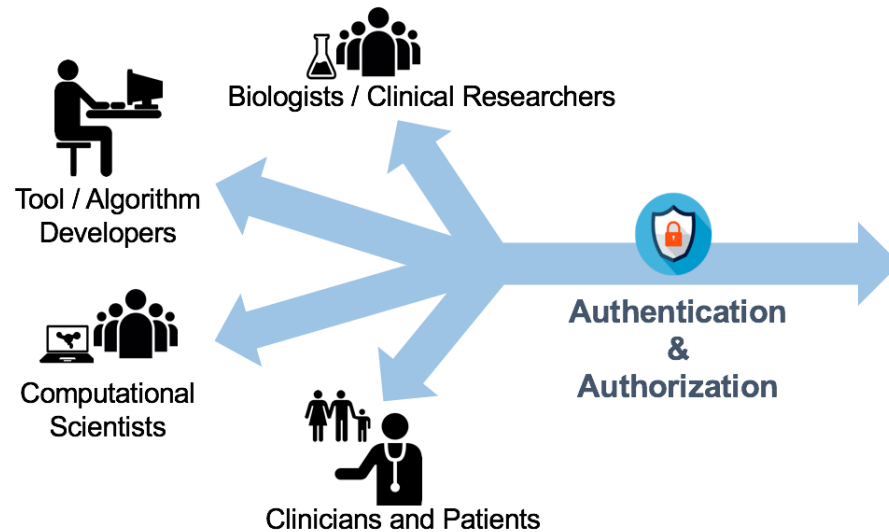
This mammogram dataset has been sourced from The Cancer Imaging Archive (TCIA) and consists of 3486 DICOM (Digital Imaging and Communications in Medicine) images. It contains normal, benign, and malignant cases with verified pathology information, allowing this to be factored into the analysis.

# TCIA data in the NCI Data Commons cloud

## The NCI Cancer Research Data Commons A virtual, expandable infrastructure

- Standardized data submission and Q/C
- Controlled vocabularies
- Harmonization by subject matter experts

- Secure data access through API or web UI
- Query across data domains
- Analytics, elastic compute, visualization



# Imaging Data Commons

Get started today! Contact us about setting up your own Google Cloud Platform Project with [free cloud credits](#)

Collections

Exploration

## RADIOLOGY

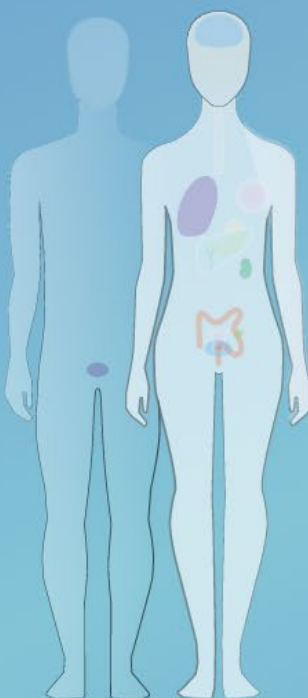


Computed Tomography (CT)

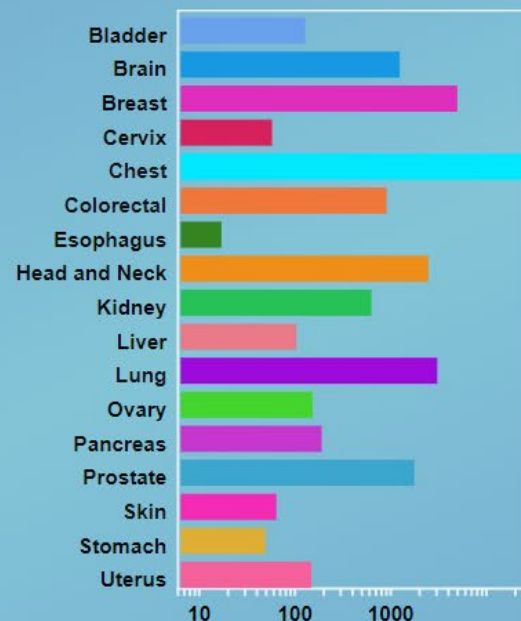
Magnetic Resonance (MR)

Positron Emission Tomography (PET)

Slide Microscopy (SM)



## Cases by Major Primary Site



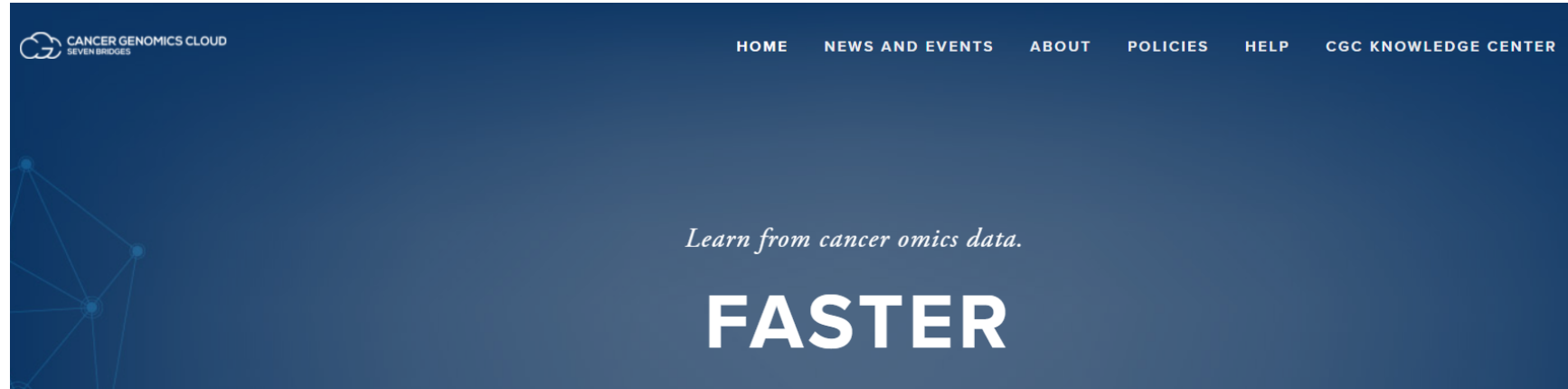
# Imaging Data Commons – Use Cases

---

- Use the IDC portal to build a cohort of cases that are needed for your use case: see [demo video - Introduction to IDC portal](#)
- Interrogate the DICOM metadata directly from a Colab notebook (or BigQuery Console): see [demo notebook - Exploration of LIDC collection](#)
- Use Google Colab notebooks to operate on the cohort: see [demo notebook - Working with IDC cohorts](#) and [the accompanying tutorial video](#)
- Use Google DataStudio to build interactive data dashboards for your cohort: see [example of using DataStudio to build a custom dashboard of IDC content](#)
- Deploy tools on a GCP Virtual Machine to support visualization and analysis of your cohort: [learn how you can launch a desktop-like VM running 3D Slicer](#)
- Develop reproducible analysis workflows operating on public data that can be easily shared with your colleagues and referenced from your paper: see [the growing list of the Colab Notebooks](#) implementing use cases relevant in cancer imaging (such as [the use of nnU-Net for segmenting lung nodules](#))

# Seven Bridges Cancer Genomics Cloud

- Combined imaging and 'omic data analyses for various NIH projects
- Bring your own tools to analyze CGC public datasets
- Add your own data to analyze alongside the public datasets using predefined analytical workflows or your own tools



The Cancer Genomics Cloud (CGC), powered by [Seven Bridges](#), is one of three [cloud resources](#) funded by the [National Cancer Institute](#) to explore the paradigm of colocalizing massive public datasets, like [The Cancer Genomics Atlas \(TCGA\)](#), [Clinical Proteomic Tumor Analysis Consortium \(CPTAC\)](#), [Therapeutically Applicable Research to Generate Effective Treatments \(TARGET\)](#) and other large omics datasets, alongside secure and scalable computational resources to analyze them. The CGC makes more than three petabytes of multi-dimensional data available immediately to authorized researchers. You can add your own data to analyze alongside the public datasets using predefined analytical workflows or your own tools. Every execution is fully reproducible, and collaborating with your team is simple and secure.



# Social Media to push new data notifications

---



Twitter: @TCIA\_News



Facebook: The Cancer Imaging Archive (page)



LinkedIn: The Cancer Imaging Archive (group)



# TCIA Acknowledgements

---

## **Fredrick National Laboratory for Cancer Research**

John Freymann  
Justin Kirby  
Brenda Fevrier-Sullivan  
Pam Angelus  
Luis Cordeiro  
Michelle Tacconelli

## **NBIA Team (NCI/FNLCR/Ellumen)**

Ulli Wagner  
Scott Gustafson  
Qinyan Pan  
Russ Rieling  
Carolyn Klinger  
Martin Lerner  
Tin Tran  
Keyvan Farahani

## **University of Arkansas**

Dr. Fred Prior  
Kirk Smith  
Lawrence Tarbox  
Dwayne Dobbins  
Jeff Tobler  
Sonya Utecht  
Julie Frund  
Bill Bennett  
Tracy Nolan  
Quasar Jarosz  
Brittney Camp  
Erica Bilello  
Claren Freeman  
Natasha Honomichl  
Geri Blake  
Diana Stockton  
Erica Bilello  
Robert Brown

## **Emory University**

Dr. Ashish Sharma  
Ryan Birmingham

## **Contractors**

John Perry  
Betty Levine

## **Imaging and Radiation Oncology Core**

Dr. Michael Knopp  
Connor Nealer