

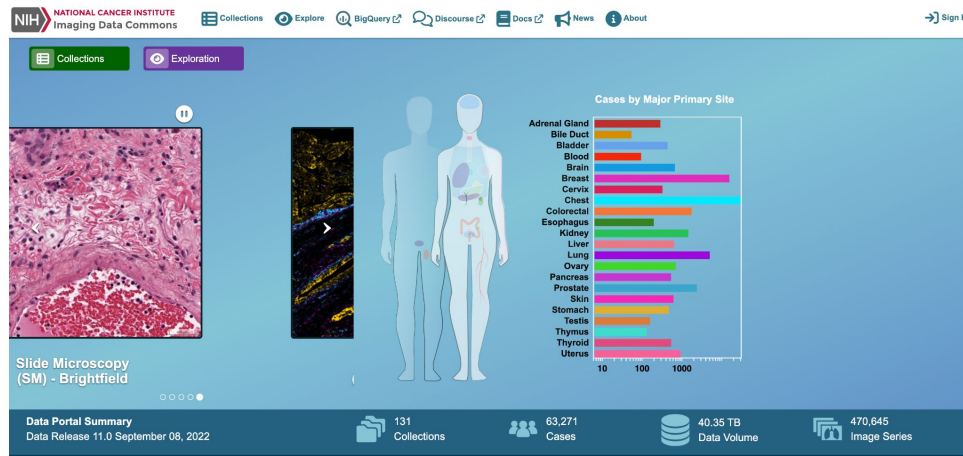
NCI Imaging Data Commons

*Andrey Fedorov, PhD, on behalf of the IDC team
Brigham and Women's Hospital, Mass General Brigham, Boston
28 November 2022*

Disclosures / disclaimer

- IDC has been funded in whole or in part with Federal funds from the NCI, NIH, under task order no. HHSN26110071 under contract no. HHSN261201500003I.
- The views presented do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

National Cancer Institute (NCI) Imaging Data Commons (IDC)



NCI Imaging Data Commons (IDC) is a **cloud-based repository** of publicly available **cancer imaging data** co-located with the analysis and exploration **tools and resources**.

Data Portal Summary
Data Release 12.0 October 11, 2022

128
Collections

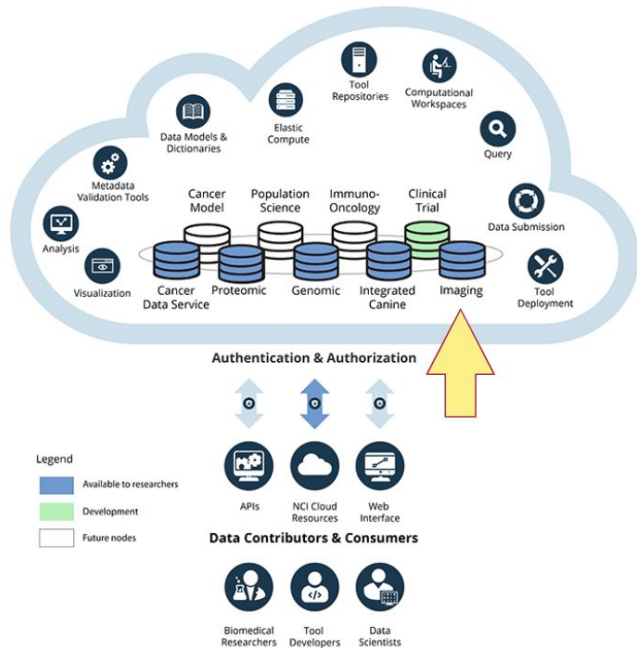
63,316
Cases

40.37 TB
Data Volume

470,850
Image Series

<https://imaging.datacommons.cancer.gov>

NCI Imaging Data Commons (IDC)



IDC is a component within the broader NCI **Cancer Research Data Commons (CRDC)** infrastructure that provides secure access to a large, comprehensive, and expanding collection of cancer research data.



Explore

Broad range of imaging data across scales, modalities, cancer types.



Subset

Extensive metadata index powered by highly scalable cloud search.



Analyze

Self-contained reproducible end-to-end cloud-based workflows.



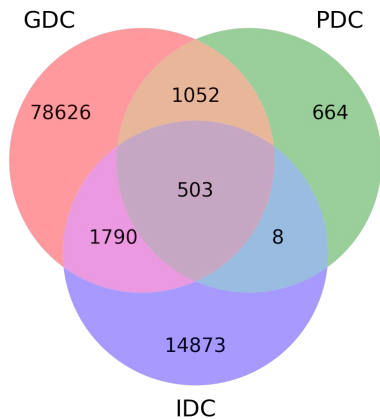
Share

Persistent data cohorts for reproducibility and transparency.



Explore

Broad range of imaging data across scales, modalities, cancer types.



NCI Imaging Data Commons in numbers

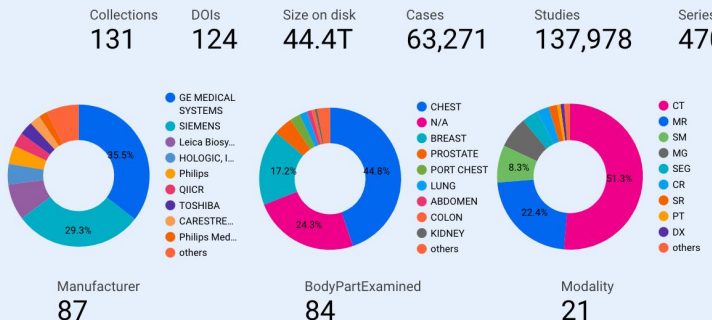


Image collections accompanied by annotations

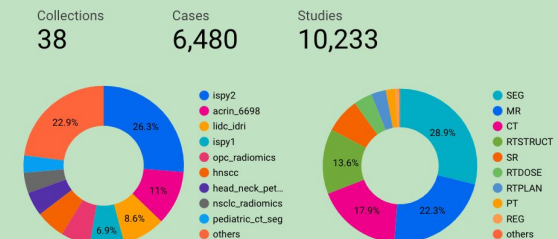
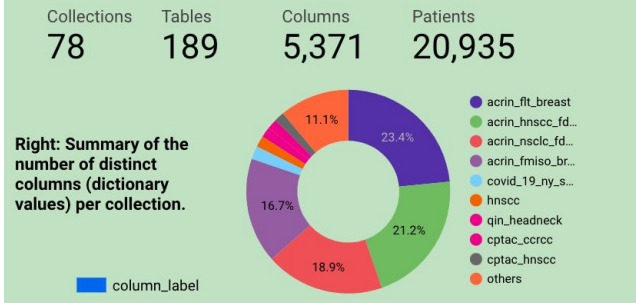
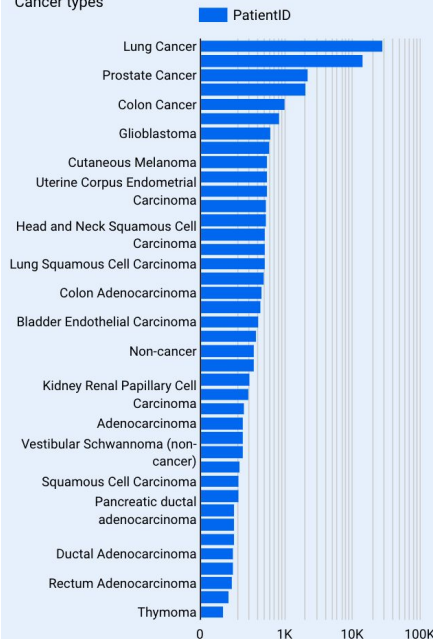


Image collections accompanied by clinical data



Cancer types





NATIONAL CANCER INSTITUTE
Imaging Data Commons

[Explore Images](#) [Collections](#) [BigQuery](#) [User Forum](#)

Cohorts

Export Manifests

Showing 1 to 10 of 35 entries Show 10 entries

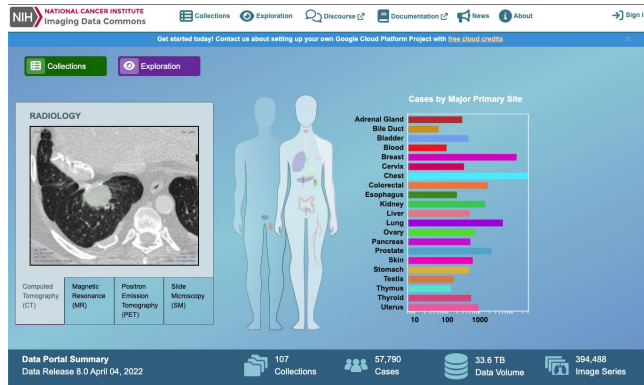
Cohort ID Name

| Case Count | Study Count | Series Count |
|------------|-------------|--------------|
|------------|-------------|--------------|

723

all of apollo

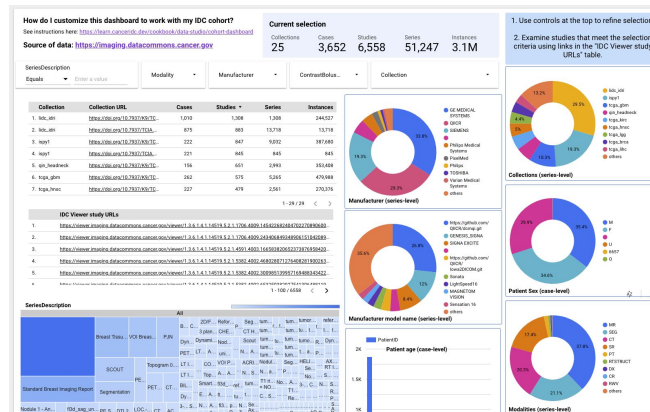
12 50 106



```

1 WITH
2   nlst_instances_per_series AS (
3     SELECT
4       StudyInstanceUID,
5       SeriesInstanceUID,
6       COUNT(DISTINCT(SOPInstanceUID)) AS num_instances,
7       COUNT(DISTINCT(ARRAY_TO_STRING(ImagePositionPatient,"/"))) AS p
8       COUNT(DISTINCT(ARRAY_TO_STRING(PixelSpacing,"/"))) AS pixel_spa
9       COUNT(DISTINCT(ARRAY_TO_STRING(ImageOrientationPatient,"/"))) A
10      MIN(SAFE_CAST(SliceThickness AS float64)) AS min_SliceThickness
11      MAX(SAFE_CAST(SliceThickness AS float64)) AS max_SliceThickness
12      MIN(SAFE_CAST(ImagePositionPatient[SAFE_OFFSET(2)] AS float64))
13      MAX(SAFE_CAST(ImagePositionPatient[SAFE_OFFSET(2)] AS float64))
14      STRING_AGG(DISTINCT(SAFE_CAST("LOCALIZER" IN UNNEST(ImageType
15 FROM
16   'bigquery-public-data.idc_current.dicom.all'

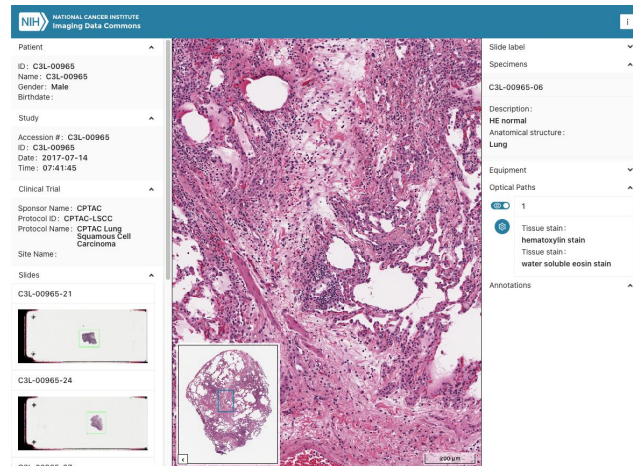
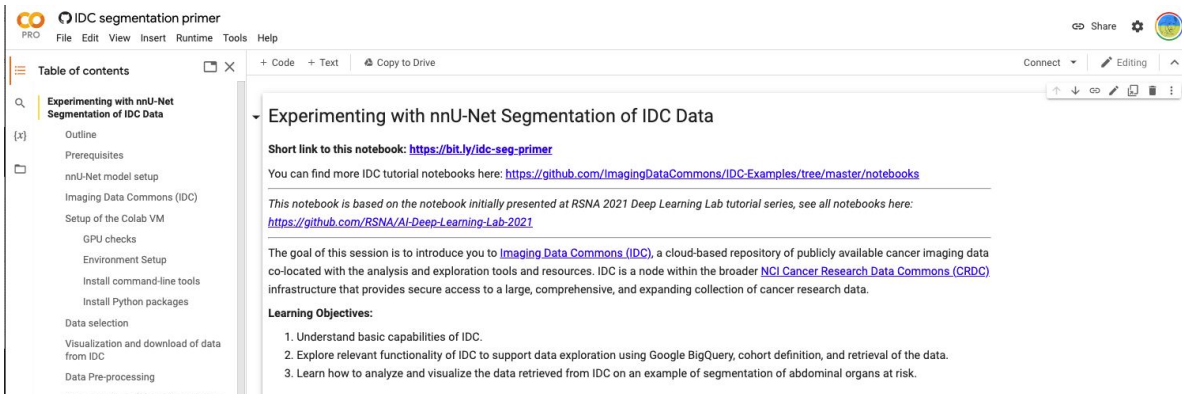
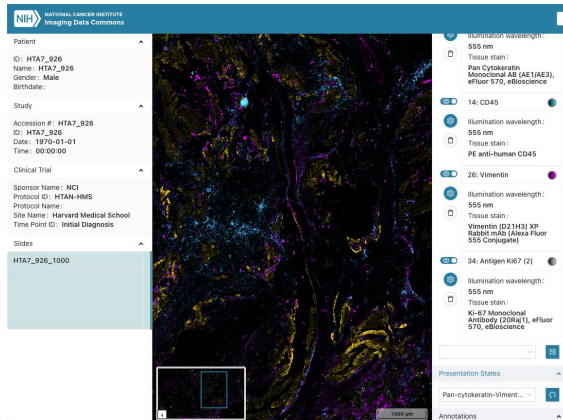
```





Analyze

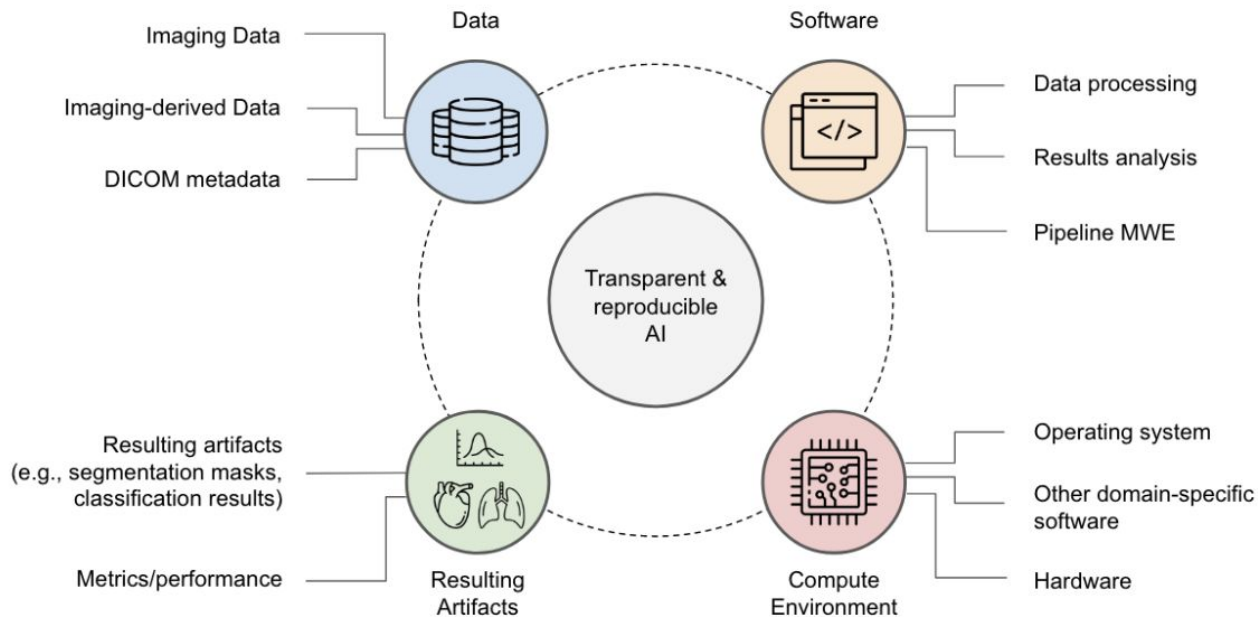
Self-contained
reproducible end-to-end
cloud-based workflows.





Share

Persistent data cohorts for reproducibility and transparency.



What you get

FOR FREE

Without downloading data or installing any software on your computer:

- Search and explore data
- Visualize images and derived data
- Define and save cohorts
- Develop/share analysis notebooks
- Find matching non-imaging data



Can I just download IDC data from the cloud and move on?

- **Yes!**

- **no out of cloud egress fees**