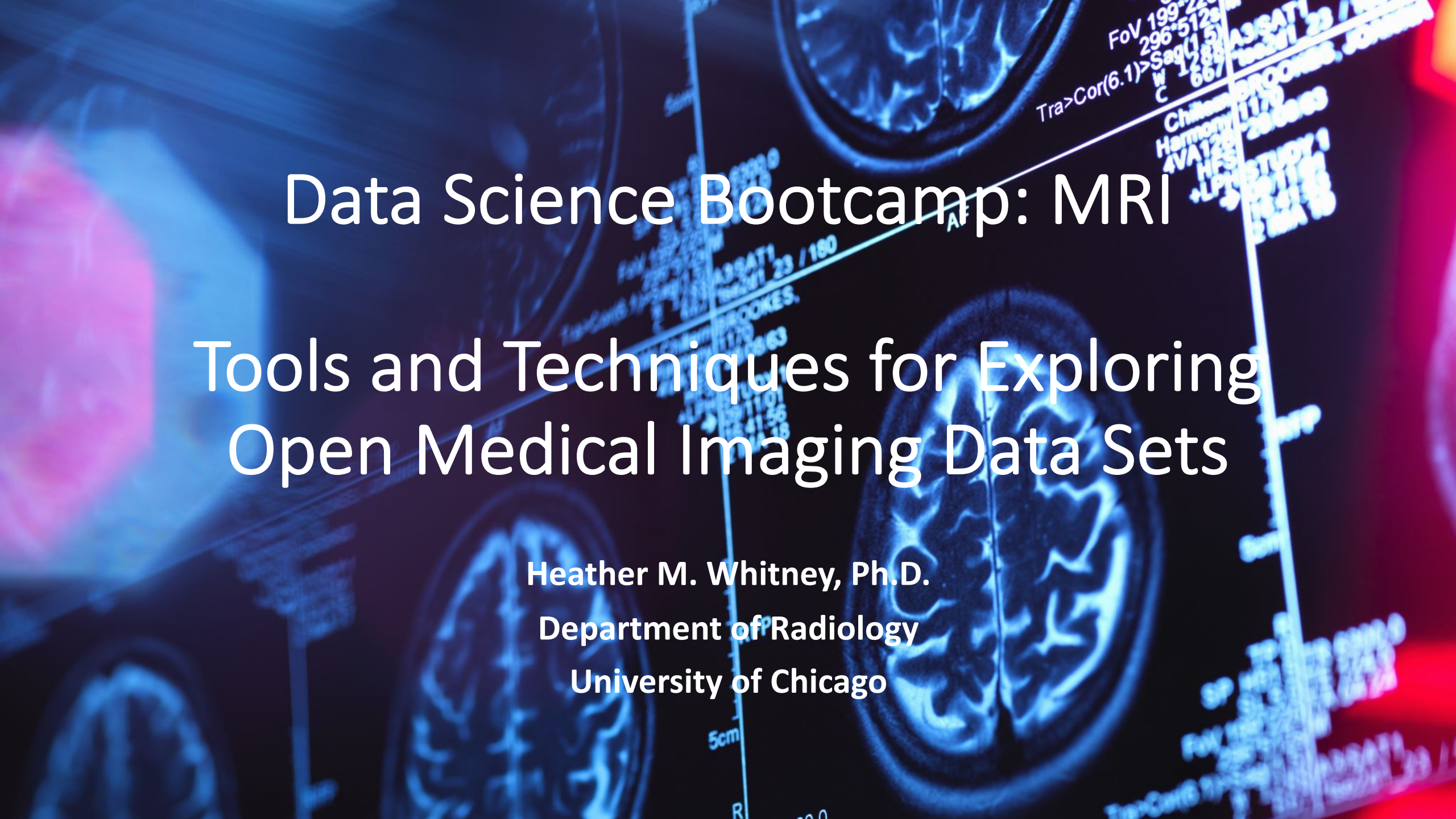


Data Science Bootcamp: MRI

Tools and Techniques for Exploring Open Medical Imaging Data Sets

Heather M. Whitney, Ph.D.

Department of Radiology

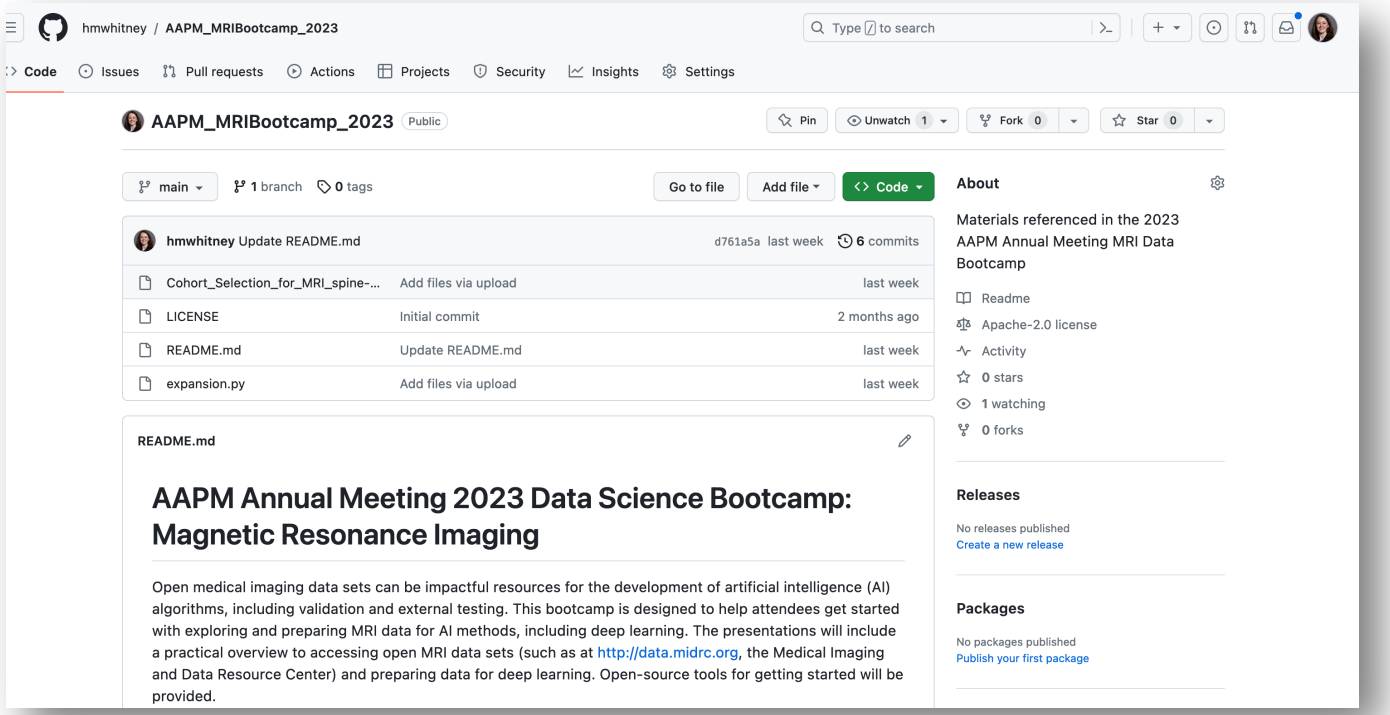
University of Chicago

Acknowledgements

Funding

The Medical Imaging Data Resource Center (MIDRC) is funded by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under contracts 75N92020C00008 and 75N92020C00021.

SCAN ME



https://github.com/hmwhitney/AAPM_MRIBootcamp_2023

Contact us

hwhitney@uchicago.edu

stephanie.harmon@nih.gov

Learning Objectives

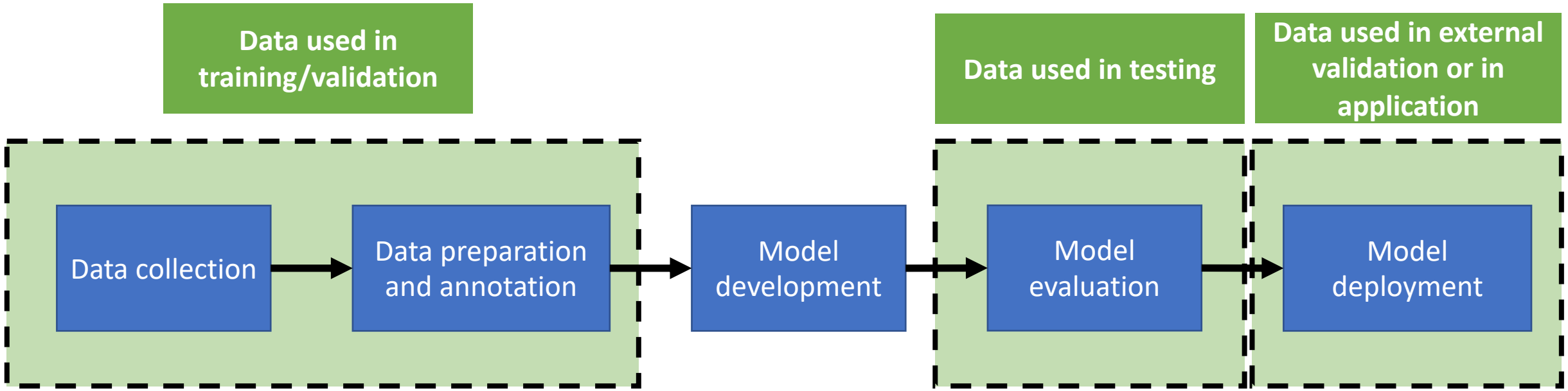
1. Become familiar with resources for accessing open MRI data sets
2. Understand how to develop a pipeline to access, download, and begin working with open MRI data sets
3. Understand techniques for pre-processing MRI data for deep learning

Open MRI Datasets

Become familiar with resources for accessing open MRI data sets

Open MRI Datasets

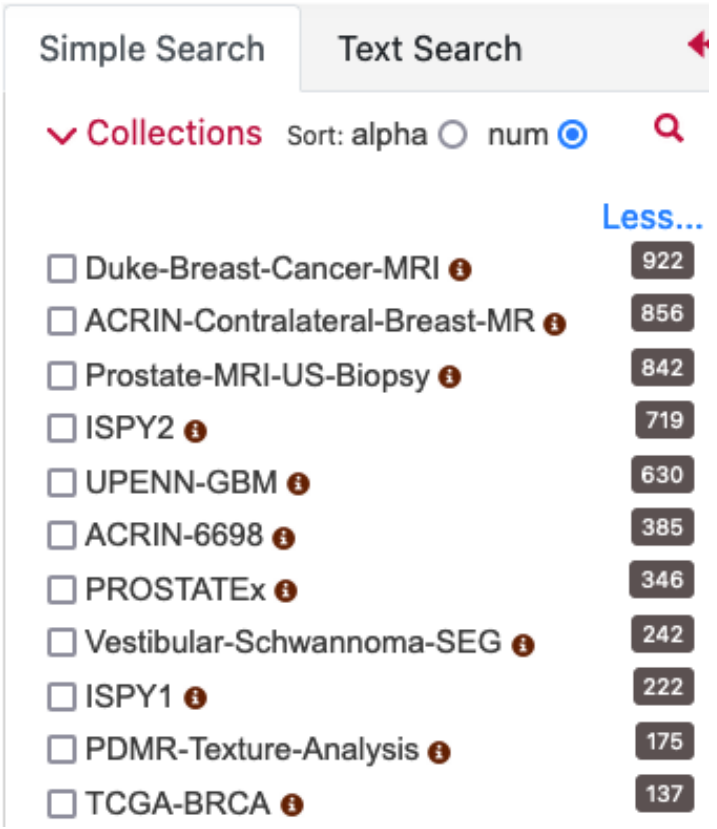
Points in an artificial intelligence medical imaging pipeline that can use open datasets



Open MRI Datasets

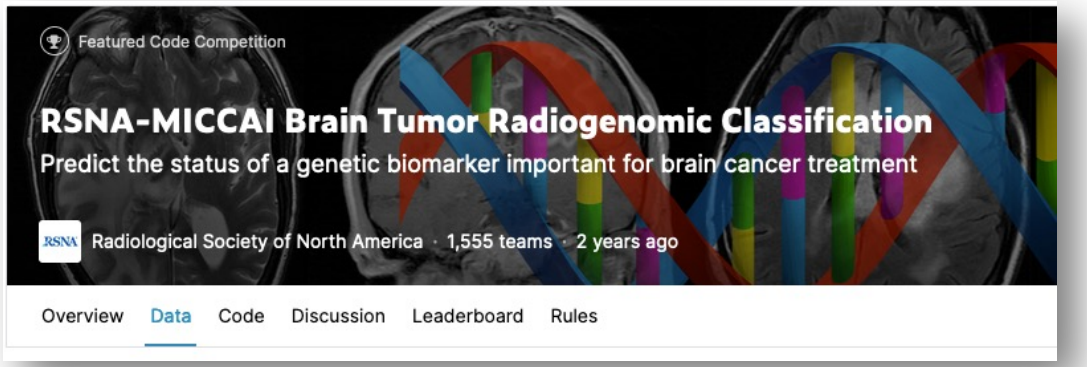
The Cancer Imaging Archive **https://www.cancerimagingarchive.net/**

NIH NATIONAL CANCER INSTITUTE
CIP Cancer Imaging Program



Open MRI Datasets

Data used in Challenges sponsored by the Radiological Society of North America



https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/brain-tumor-ai-challenge-2021

- Fluid Attenuated Inversion Recovery (FLAIR)
- T1-weighted pre-contrast (T1w)
- T1-weighted post-contrast (T1Gd)
- T2-weighted (T2)

Open MRI Datasets

Data used in Challenges sponsored by the AAPM

Hosted at TCIA

The Medical Imaging and Data Resource Center (MIDRC)

Medical Imaging Community Response to the COVID-19 Pandemic

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

RSNA Radiological Society of North America | AAPM American Association of Physicists in Medicine | ACR American College of Radiology | The University of Chicago | GEN3 Data Commons | NIH National Institute of Biomedical Imaging and Bioengineering

MIDRC is:

- A curated image and data commons
- A means by which researchers can address topics no single image archive could yield independently
- A partnership of the AAPM, ACR and RSNA, supported by NIBIB and hosted at University of Chicago

There are two scientific components of MIDRC:

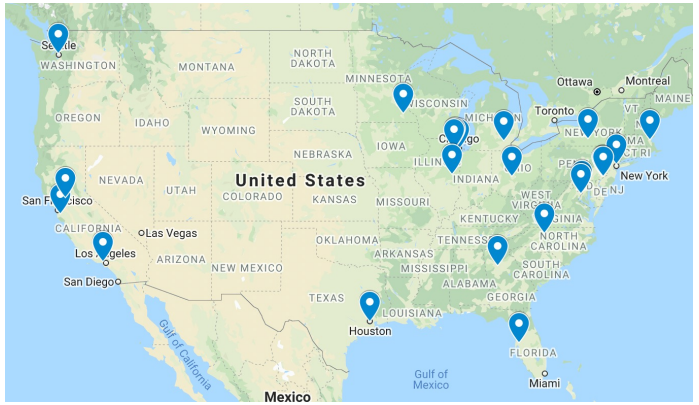
1. *Open Discovery Data Commons*

   Creation, testing, quality assurance, and data connectivity

2. *Machine Intelligence Computational Capabilities*

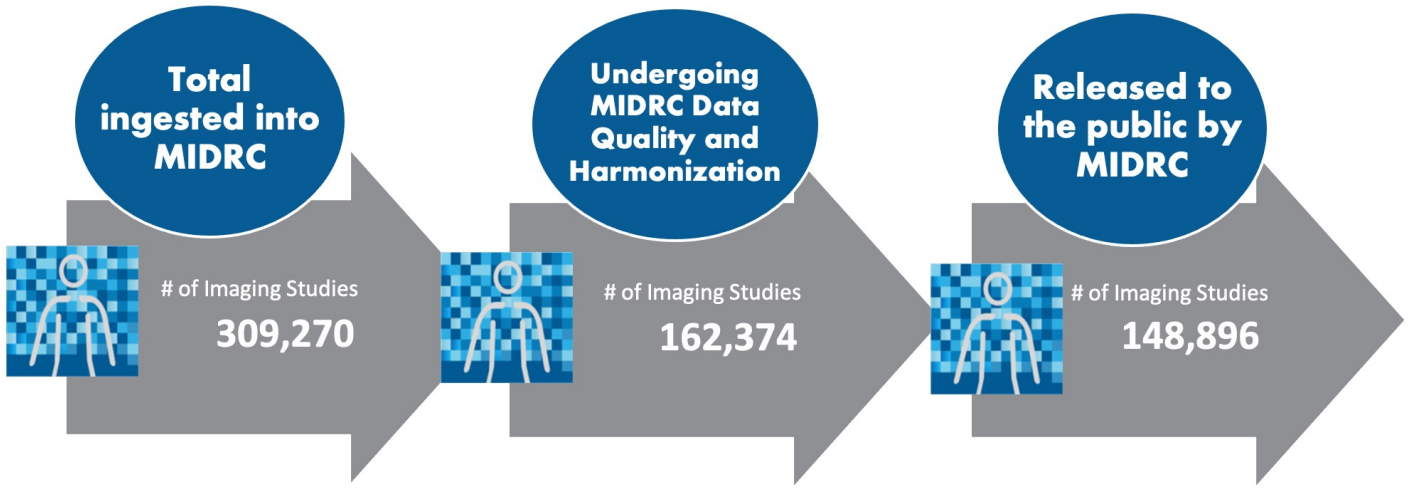
   Clinically relevant algorithms and software tools

Established August 2020
Supported by NIBIB



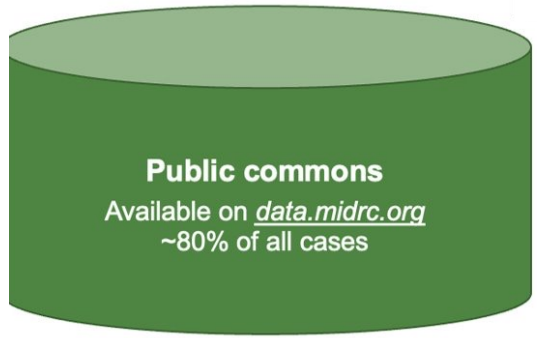
A collaboration of 24+ institutions & 100+ investigators from academia, community practices, FDA, and others

The Medical Imaging and Data Resource Center (MIDRC)



GEN3 DATA COMMONS

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

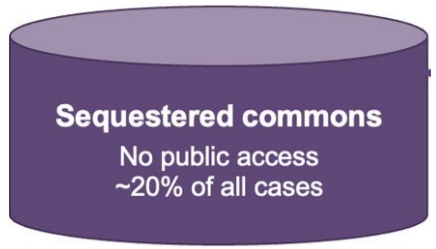


Public facing DATA

- AI/ML ready
- FAIR
- Trustworthy
- Representative
- Searchable

- De-identification
- Curation
- Quality control
- Annotation
- Harmonization

(incl. LOINC mapping)



Sequestered DATA

- Not publicly available
- Real-world testing of algorithm performance
- Translation / support of regulatory submission

58,779
Cases

13.03TB
Total size Published

Curation and harmonization of DICOM images and metadata

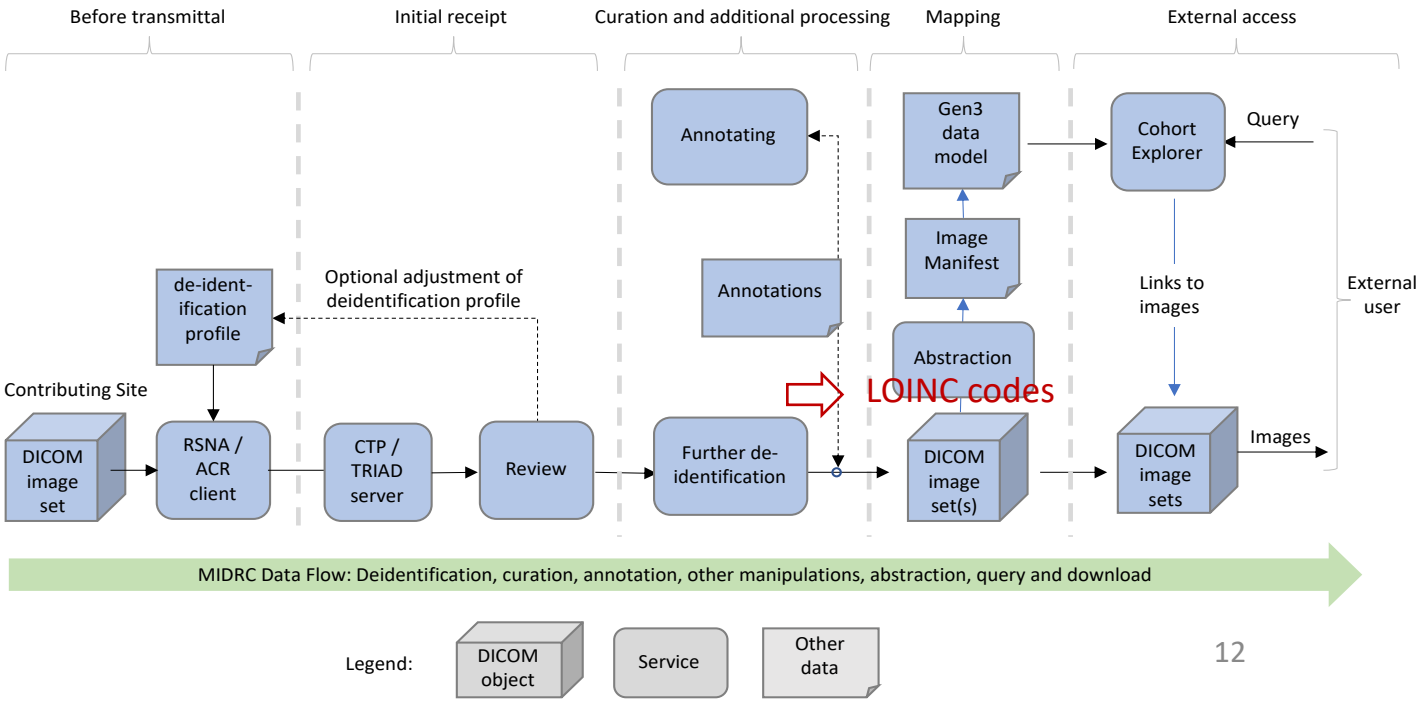
The quality of data collection is important

Subset of a public *(not MIDRC)* COVID-19 DICOM chest x-ray image collection showing variations in image quality and view directions and body part (i.e., the knee image near center)

The MIDRC data ingestion process includes

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

- Multiple DICOM study descriptions from incoming MIDRC data
- Matched to a unique LOINC code
- Enables searching on Gen3 user portal for cohort building



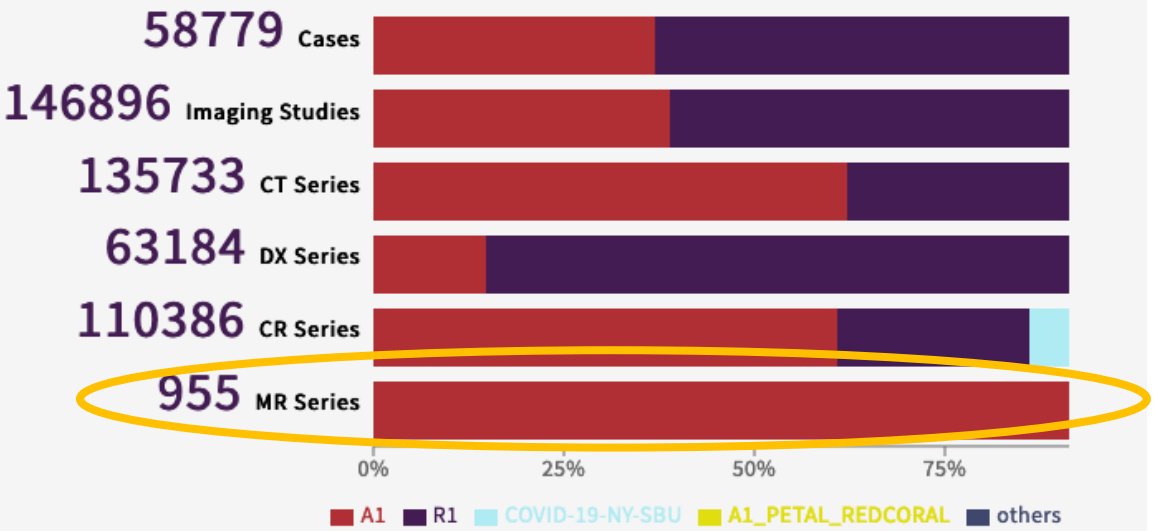
Open MRI Data available at MIDRC

http://data.midrc.org

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

MIDRC Data Commons

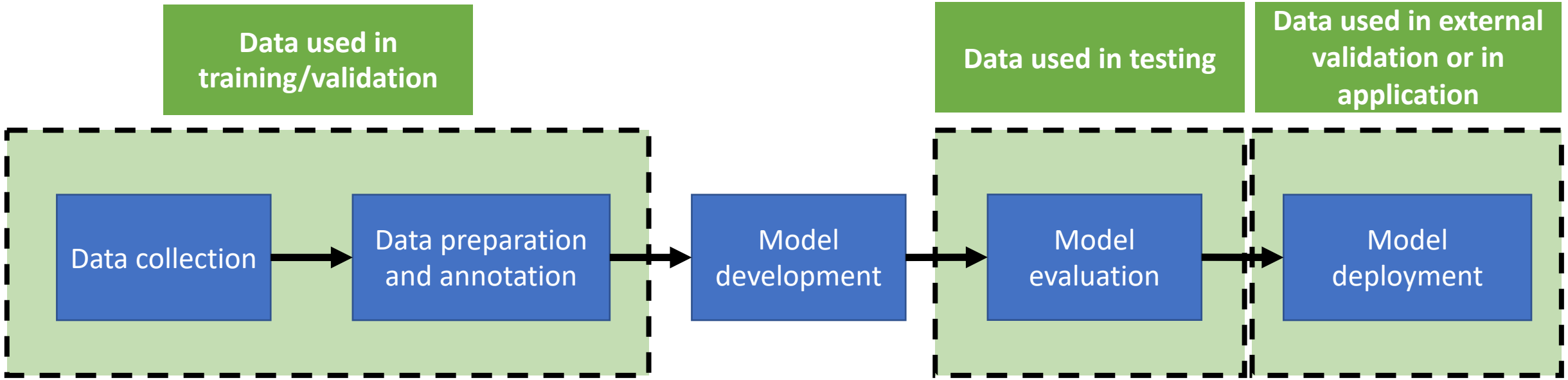
The Medical Imaging & Data Resource Center (MIDRC) Data Commons supports the management, analysis and sharing of medical imaging data for the improvement of patient outcomes. The data in MIDRC are open access in order to foster machine learning innovation through data sharing and include in addition to imaging files, patient demographic data, COVID-19 test results and other clinical data, harmonized study descriptions utilizing the LOINC playbook, and image DICOM tags for purposes of data filtering and selecting cohorts for analysis.



Developing a Pipeline

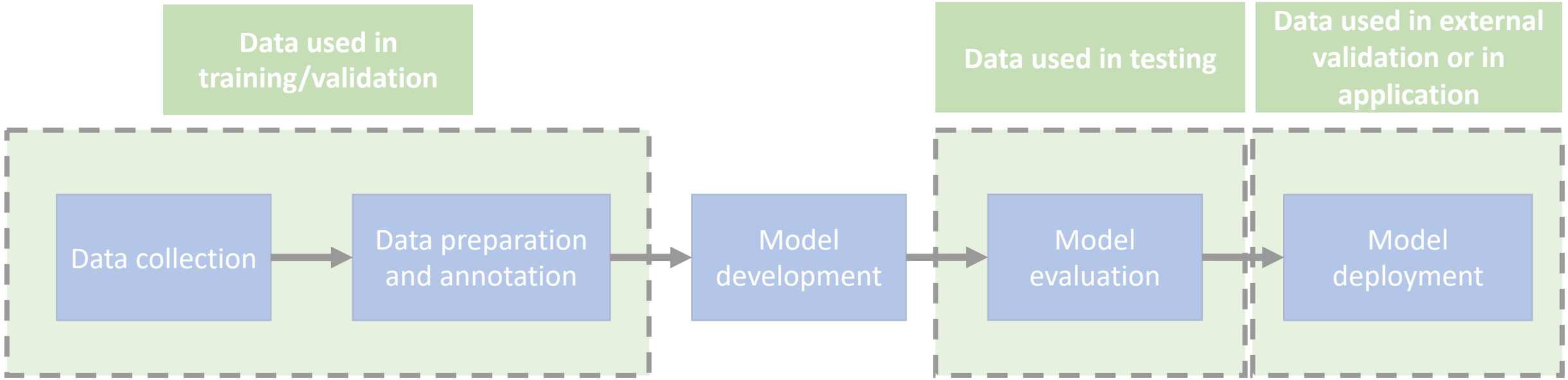
Understand how to develop a pipeline to access, download, and begin working with open MRI data sets

Developing a pipeline



Developing a pipeline

Access



Developing a pipeline

Gen3

Gen3 is a data platform for building data commons and data ecosystems.

The Gen3 platform consists of open-source software

The AnVIL

NIH National Heart, Lung, and Blood Institute | BioData CATALYST

Biomedical Research Hub Powered by Gen3

BloodPAC BLOOD PROFILING ATLAS IN CANCER

CANINE Data Commons

CHICAGOLAND COVID-19 COMMONS

NIH NATIONAL CANCER INSTITUTE Cancer Research Data Commons

Environmental Data Commons

GenoMEL the Melanoma Genetics Consortium

NIH HEAL INITIATIVE

BDGC

NIH HEAL INITIATIVE | Justice Community Opioid Innovation Network (JCOIN)

Gabriella Miller Kids First Pediatric Research Program Data Resource Center

MIDRC MEDICAL IMAGING AND DATA RESOURCE CENTER

NIH National Institute of Allergy and Infectious Diseases | AccessClinicalData@NIAID

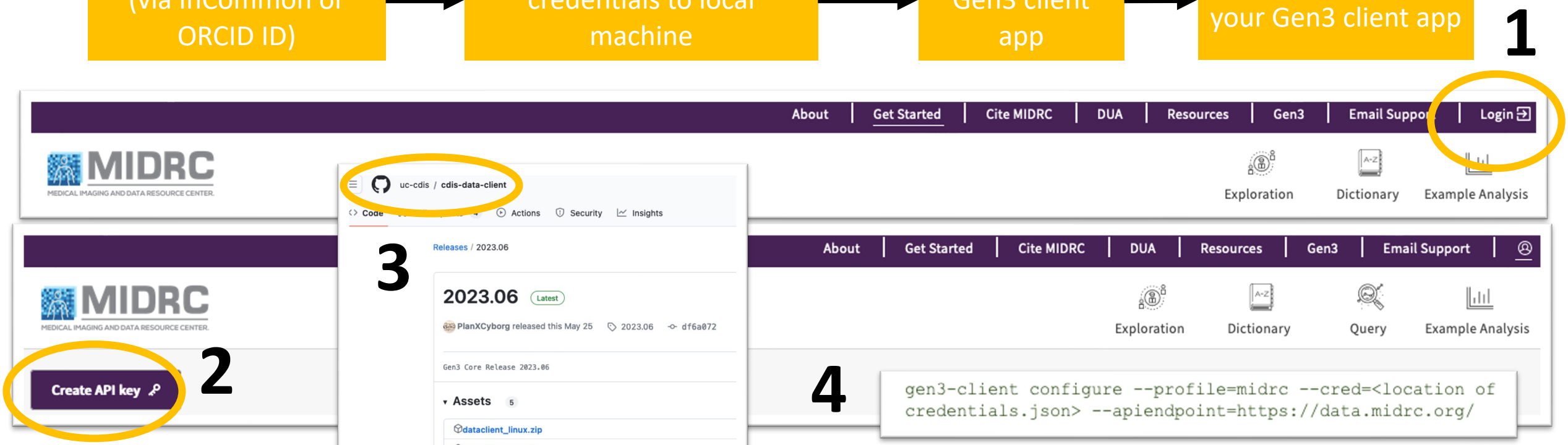
Veterans Affairs Data Commons

VPO Veterans Precision Oncology Data Commons

Developing a pipeline

Access

1. Create an account (via InCommon or ORCID ID)
2. Create an API key and save credentials to local machine
3. Install the Gen3 client app
4. Link your API key to your Gen3 client app



Developing a pipeline

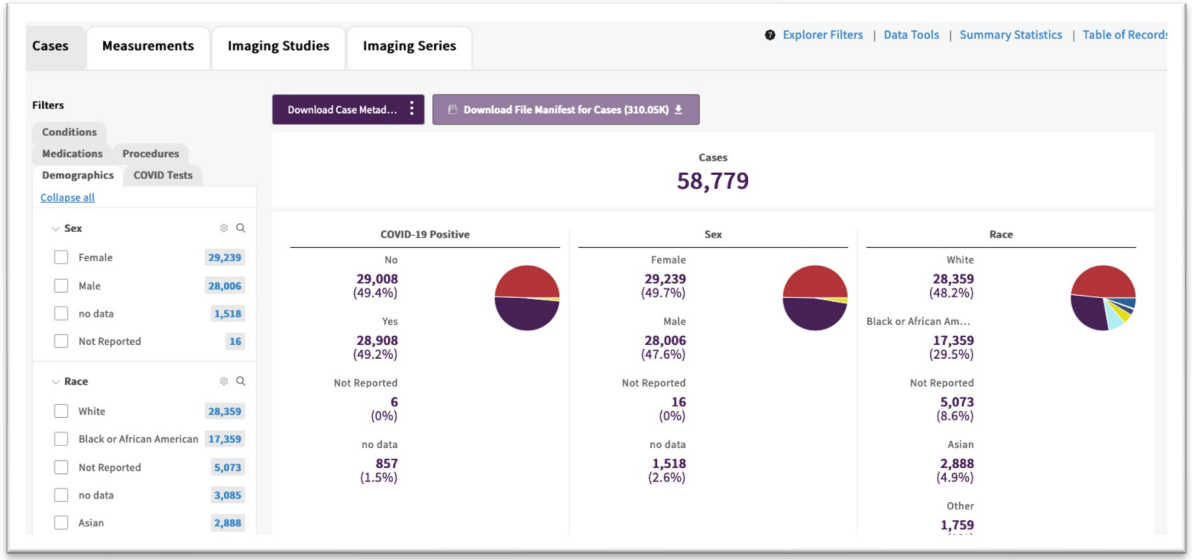
Cohort building

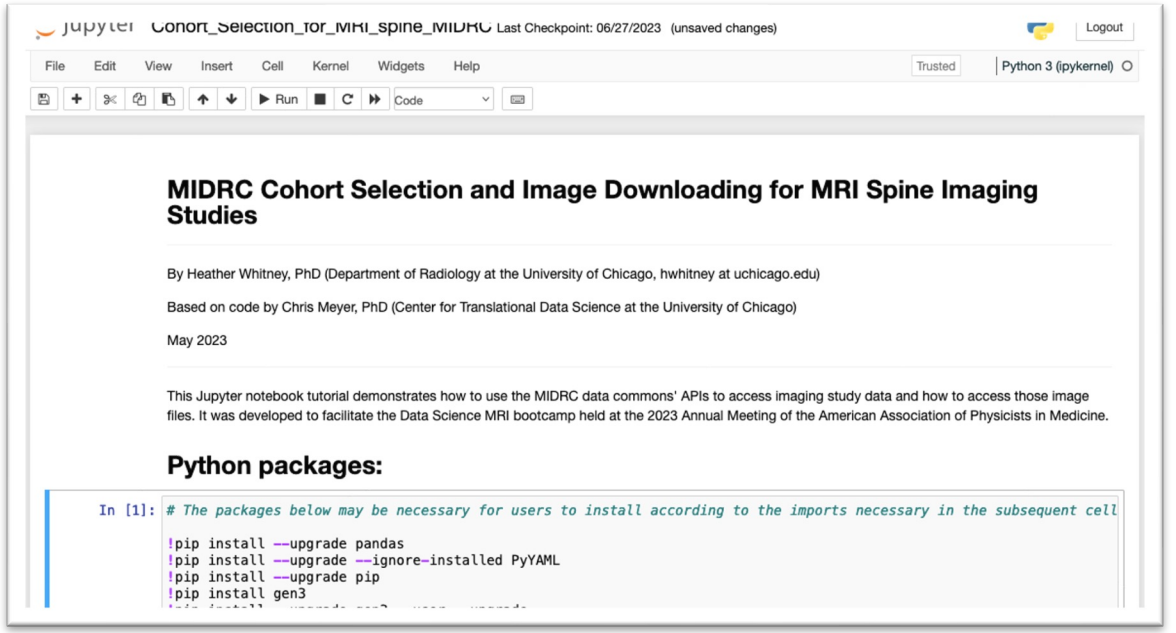
Create a cohort

How do you want to create your cohort?

Use the Exploration Tool to create your cohort

Use a Jupyter notebook to create your cohort

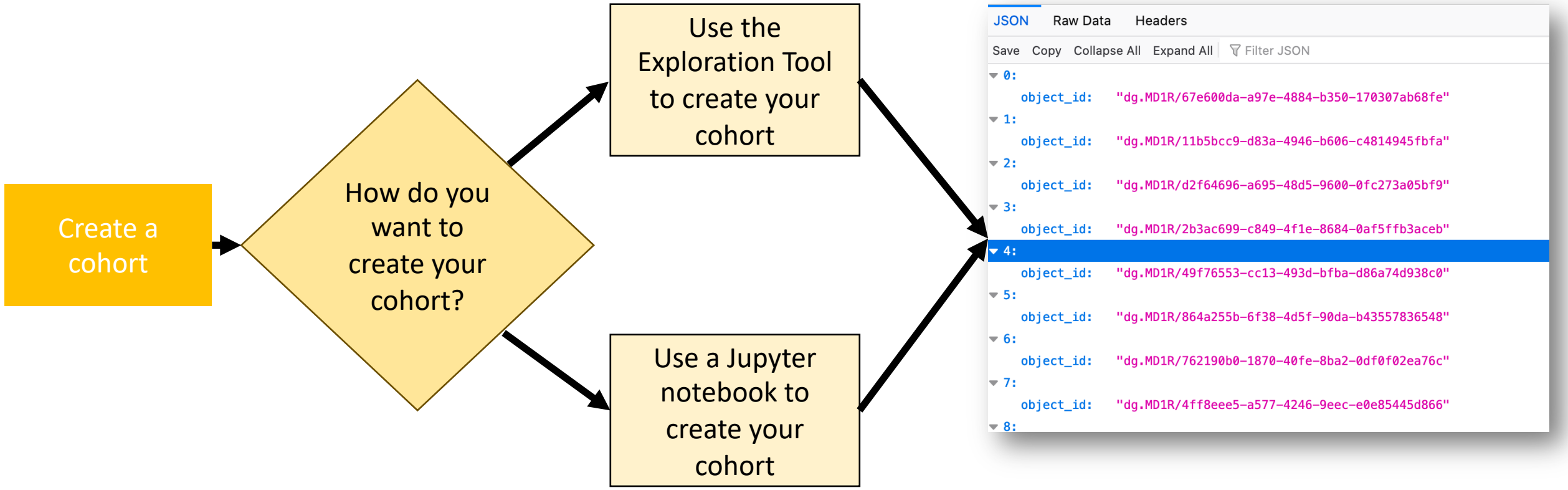




Developing a pipeline

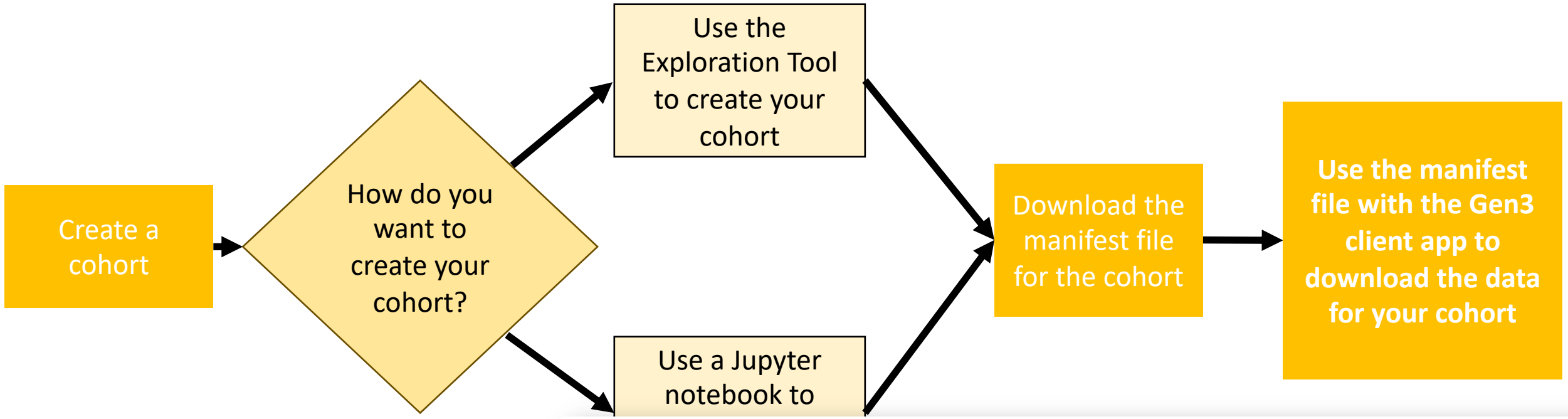
Developing a pipeline

Download



Developing a pipeline

Download




```
In [15]: # To download all files in the manifest, use the "gen3 drs-pull manifest" command
download_dir = "{}/images".format(demo_dir)

if not os.path.exists(download_dir):
    os.makedirs(download_dir)

cmd = "gen3 --auth {} --endpoint data.midrc.org drs-pull manifest {} {}".format(cred, mani_name, download_dir)
print(cmd)


## Option 1: To monitor the progress in real-time, you can copy the command generated below
## and run in your terminal instead of from this Jupyter Notebook.

gen3 --auth /Users/heatherwhitney/Documents/MRIbootcamp/credentials.json --endpoint data.midrc.org drs-pull manifes
t MRImanifest20230704.json /Users/heatherwhitney/Documents/MRIbootcamp/images
```

Begin working with data