

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

hmwhitney / AAPM_MRIBootcamp_2023

Type ⌘ to search

Code Issues Pull requests Actions Projects Security Insights Settings

AAPM_MRIBootcamp_2023 Public

main · 1 branch · 0 tags

Go to file Add file Code

About Materials referenced in the 2023 AAPM Annual Meeting MRI Data Bootcamp

Readme Apache-2.0 license

Activity 0 stars

1 watching

0 forks

LICENSE

README.md

expansion.py

README.md

**AAPM Annual Meeting 2023 Data Science Bootcamp:
Magnetic Resonance Imaging**

Open medical imaging data sets can be impactful resources for the development of artificial intelligence (AI) algorithms, including validation and external testing. This bootcamp is designed to help attendees get started with exploring and preparing MRI data for AI methods, including deep learning. The presentations will include a practical overview to accessing open MRI data sets (such as at <http://data.midrc.org>, the Medical Imaging and Data Resource Center) and preparing data for deep learning. Open-source tools for getting started will be provided.

Contact us

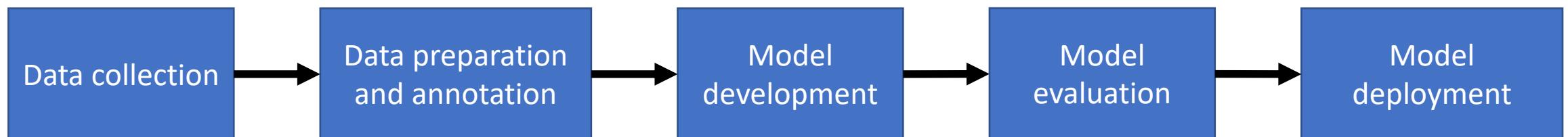
hwhitney@uchicago.edu

stephanie.harmon@nih.gov

https://github.com/hmwhitney/AAPM_MRIBootcamp_2023

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

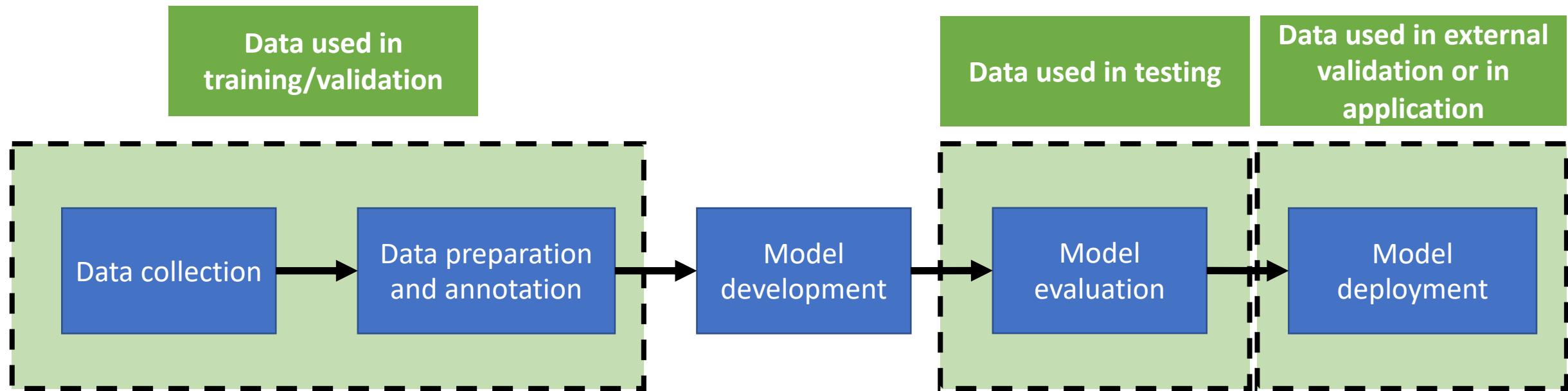


Open MRI Datasets

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/>



| Simple Search | Text Search | ◀ |
|--|--|---------|
| ▼ Collections | Sort: alpha <input type="radio"/> num <input checked="" type="radio"/> | 🔍 |
| | | Less... |
| <input type="checkbox"/> Duke-Breast-Cancer-MRI ⓘ | 922 | |
| <input type="checkbox"/> ACRIN-Contralateral-Breast-MR ⓘ | 856 | |
| <input type="checkbox"/> Prostate-MRI-US-Biopsy ⓘ | 842 | |
| <input type="checkbox"/> ISPY2 ⓘ | 719 | |
| <input type="checkbox"/> UPENN-GBM ⓘ | 630 | |
| <input type="checkbox"/> ACRIN-6698 ⓘ | 385 | |
| <input type="checkbox"/> PROSTATEx ⓘ | 346 | |
| <input type="checkbox"/> Vestibular-Schwannoma-SEG ⓘ | 242 | |
| <input type="checkbox"/> ISPY1 ⓘ | 222 | |
| <input type="checkbox"/> PDMR-Texture-Analysis ⓘ | 175 | |
| <input type="checkbox"/> TCGA-BRCA ⓘ | 137 | |

Open MRI Datasets

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



A screenshot of a competition page from the Radiological Society of North America (RSNA). The title is "RSNA-MICCAI Brain Tumor Radiogenomic Classification". The subtitle is "Predict the status of a genetic biomarker important for brain cancer treatment". It shows a grayscale MRI scan of a brain with overlaid colored regions. A DNA helix graphic is visible in the background. The RSNA logo is in the bottom left. Below the main image, there are navigation links: Overview, Data (underlined), Code, Discussion, Leaderboard, and Rules.

<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



Social media icons: Twitter, Facebook, LinkedIn, Email.

Navigation: Submit Your Data, Access The Data, Help.

The Cancer Imaging Archive logo.

Confluence Spaces: The Cancer Imaging Archive (TCIA) Public Access.

Pages / Wiki / Collections.

AAPM RT-MAC Grand Challenge 2019 (AAPM-RT-MAC)

Created by Kirk Smith, last modified by Tracy Nolan on Feb 08, 2023.

Summary

This data set was provided in association with a challenge competition and related conference session conducted at the AAPM 2019 Annual Meeting.

MRI is popular in radiation oncology because of its excellent imaging quality of soft tissue and tumor. With the advent of MR-Linac and MR-guided radiation therapy, there is a trend toward a MR-based radiation treatment planning. Contouring is an important task in modern radiation treatment planning and frequently introduces uncertainties in radiation therapy due to observer variabilities. Auto-segmentation has been demonstrated as an effective approach to reduce this uncertainty. The overall objective of this grand challenge is to provide a platform for comparison of various auto-segmentation algorithms when they are used to delineate organs at risk (OARs) or tumors from MR images for head and neck patients for radiation treatment planning. The results will provide an indication of the performances achieved by various auto-segmentation algorithms and can be used to guide the selection of these algorithms for clinic use if desirable.

The data for this challenge contains a total of 55 MRI cases, each from a single examination from a distinct patient, with each case consisting of a T2-weighted MRI images in DICOM format. The MRI scanning protocol was designed for radiation treatment simulation. Thirty-one of these will be provided as training cases, with the parotid glands, submandibular glands, level 2 and level 3 lymph nodes contoured. The images and contours were acquired from MD Anderson Cancer Center.

More details on accessing the various challenge subsets (training, off-site test, and live test) can be found on the Detailed Description tab below.

Hosted at TCIA

The Medical Imaging and Data Resource Center (MIDRC)

Medical Imaging Community Response
to the COVID-19 Pandemic



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

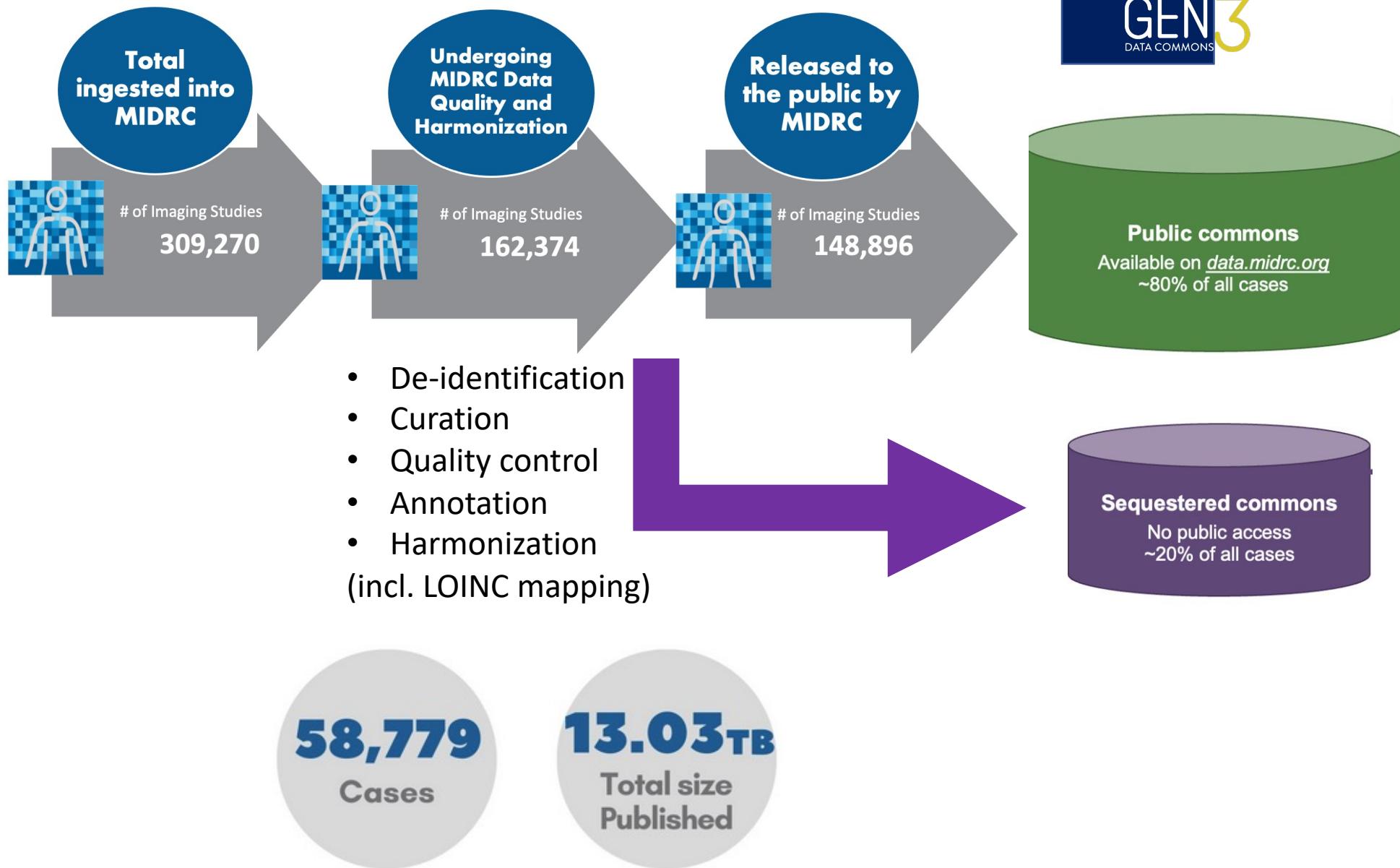


Established August 2020
Supported by NIBIB



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

The Medical Imaging and Data Resource Center (MIDRC)



Public facing DATA

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

Sequestered DATA

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

Curation and harmonization of DICOM images and metadata

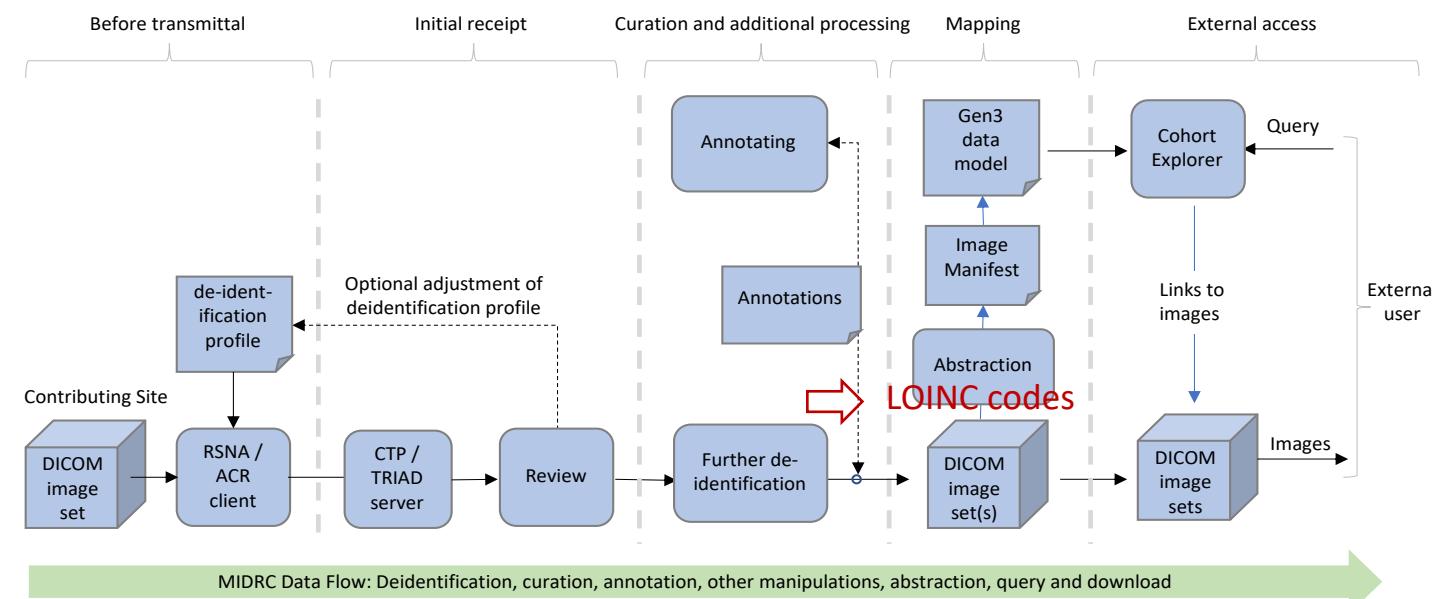
The quality of data collection is important



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

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



Legend:



Service



12

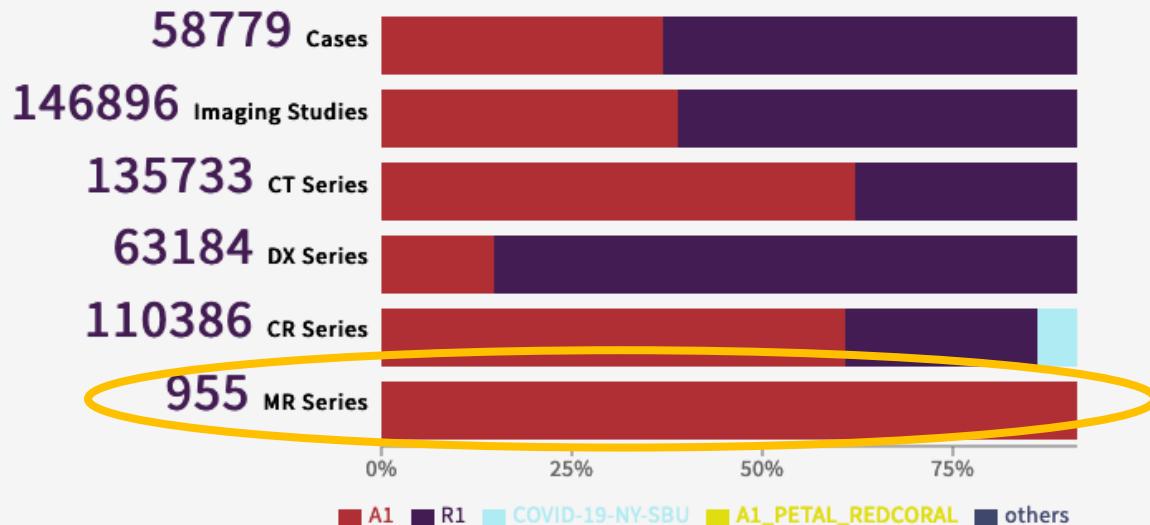
Open MRI Data available at MIDRC

<http://data.midrc.org>



MIDRC Data Commons

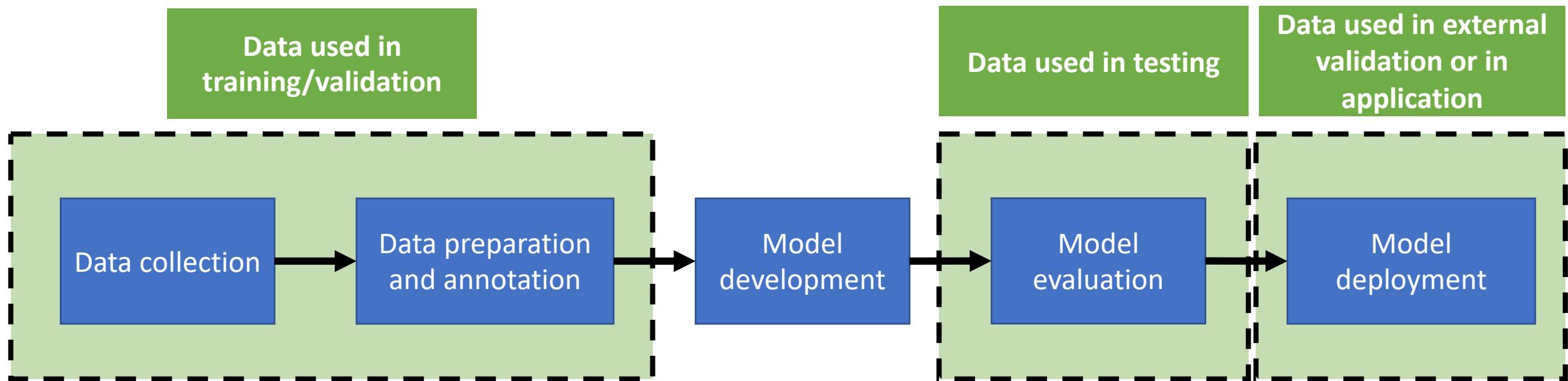
The Medical Imaging & Data Resource Center (MIDRC) Data Commons supports the management, analysis and sharing of medical imaging data for the improvement of patient outcomes. The data in MIDRC are open access in order to foster machine learning innovation through data sharing and include in addition 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



Data used in training/validation

Data used in testing

Data used in external validation or in application



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 BioData CATALYST | Biomedical Research Hub
Powered by Gen3 | BloodPAC
CANINE Data Commons | CHICAGOLAND COVID-19 COMMONS | NATIONAL CANCER INSTITUTE Cancer Research Data Commons
CCC Environmental Data Commons | GenoMEL the Melanoma Genetics Consortium | NIH HEAL INITIATIVE | BDGC | MIDRC
NIH HEAL INITIATIVE Justice Community Opioid Innovation Network (JCOIN) | Kids First Genetic Research Program Data Resource Center | VPO Veterans Affairs Data Commons | National Institute of Allergy and Infectious Diseases AccessClinicalData@NIAID | 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

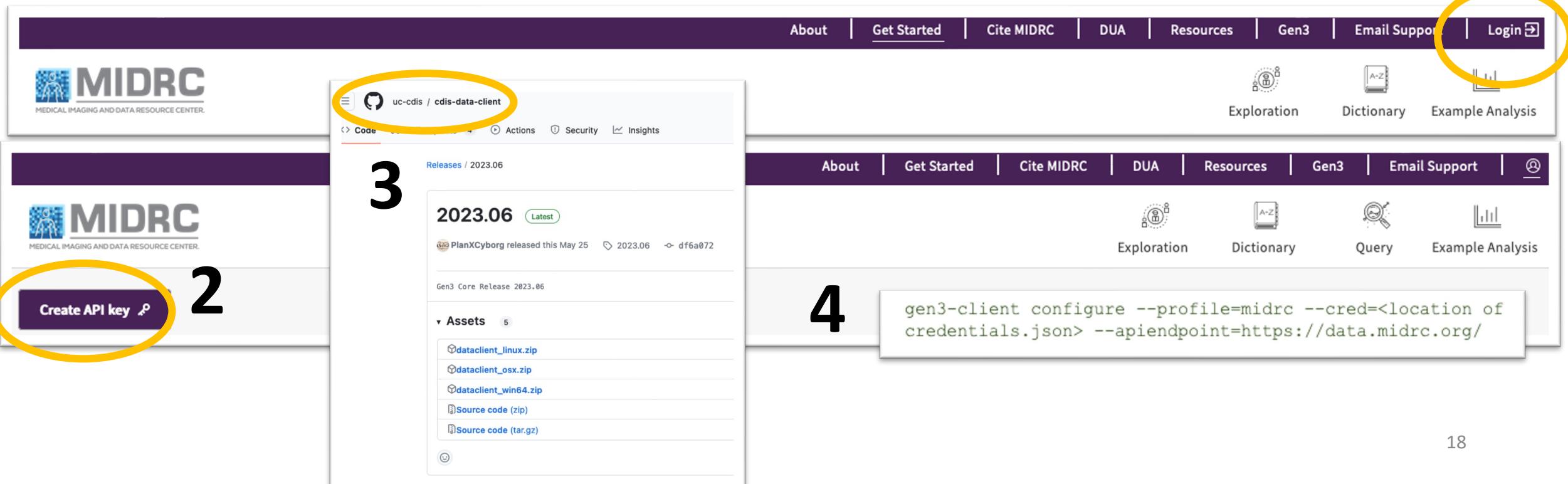
3

Install the
Gen3 client
app

4

Link your API key to
your Gen3 client app

1



Developing a pipeline

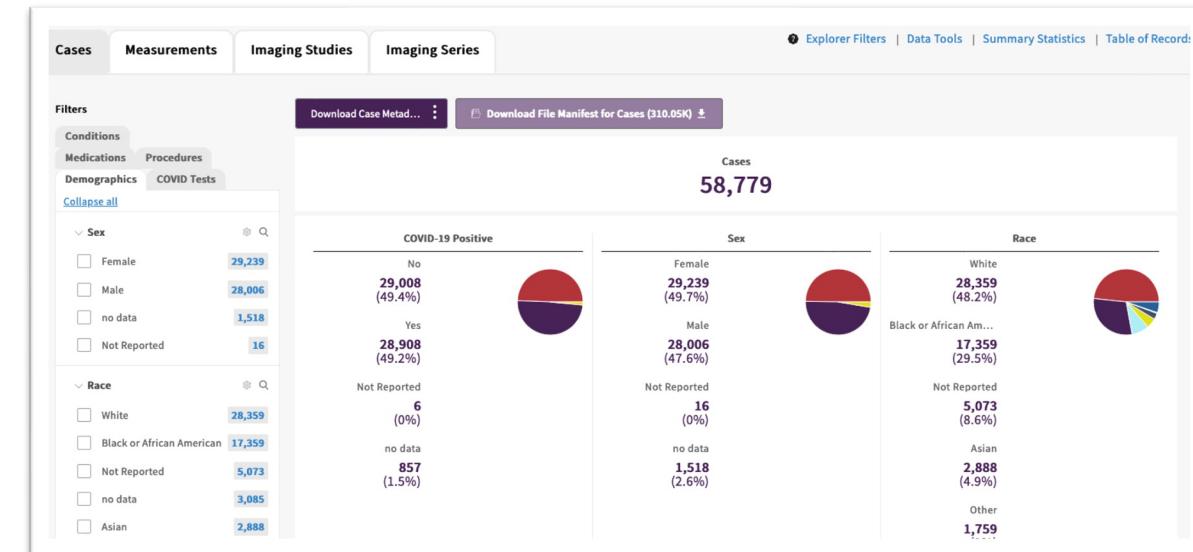
Cohort building

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



Jupyter Notebook titled 'Cohort_Selection_for_MRI_spine_MIDRC' Last Checkpoint: 06/27/2023 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3 (ipykernel) Logout

MIDRC Cohort Selection and Image Downloading for MRI Spine Imaging Studies

By Heather Whitney, PhD (Department of Radiology at the University of Chicago, hwhitney at uchicago.edu)
Based on code by Chris Meyer, PhD (Center for Translational Data Science at the University of Chicago)
May 2023

This Jupyter notebook tutorial demonstrates how to use the MIDRC data commons' APIs to access imaging study data and how to access those image files. It was developed to facilitate the Data Science MRI bootcamp held at the 2023 Annual Meeting of the American Association of Physicists in Medicine.

Python packages:

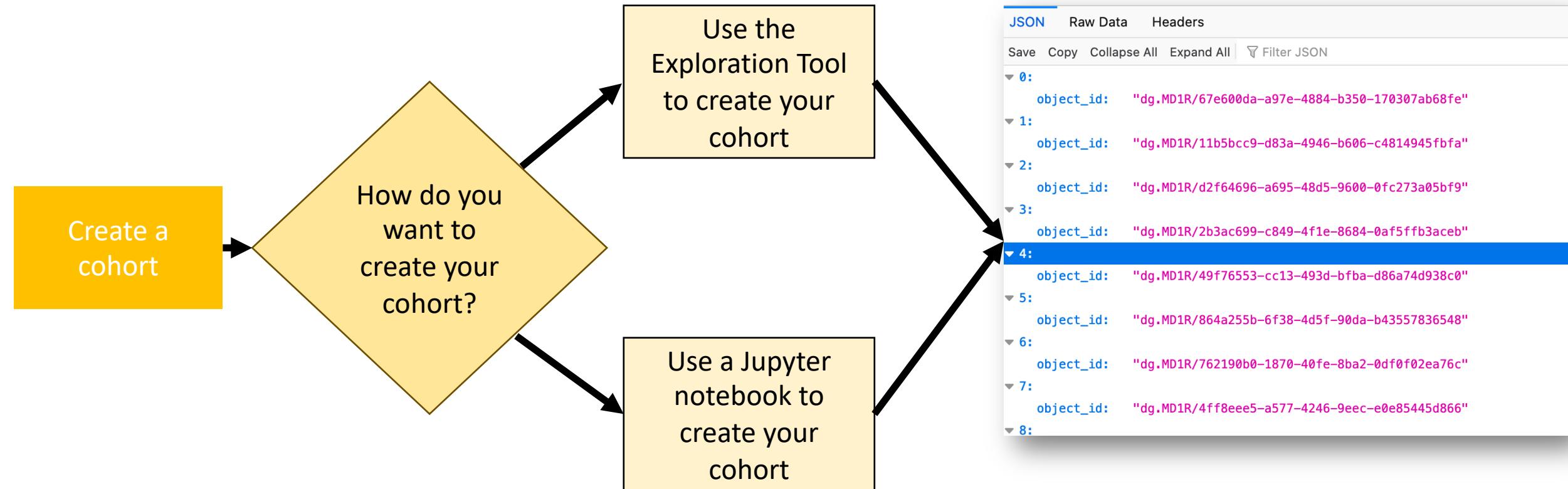
```
In [1]: # The packages below may be necessary for users to install according to the imports necessary in the subsequent cell
!pip install --upgrade pandas
!pip install --upgrade --ignore-installed PyYAML
!pip install --upgrade pip
!pip install gen3
```

Developing a pipeline



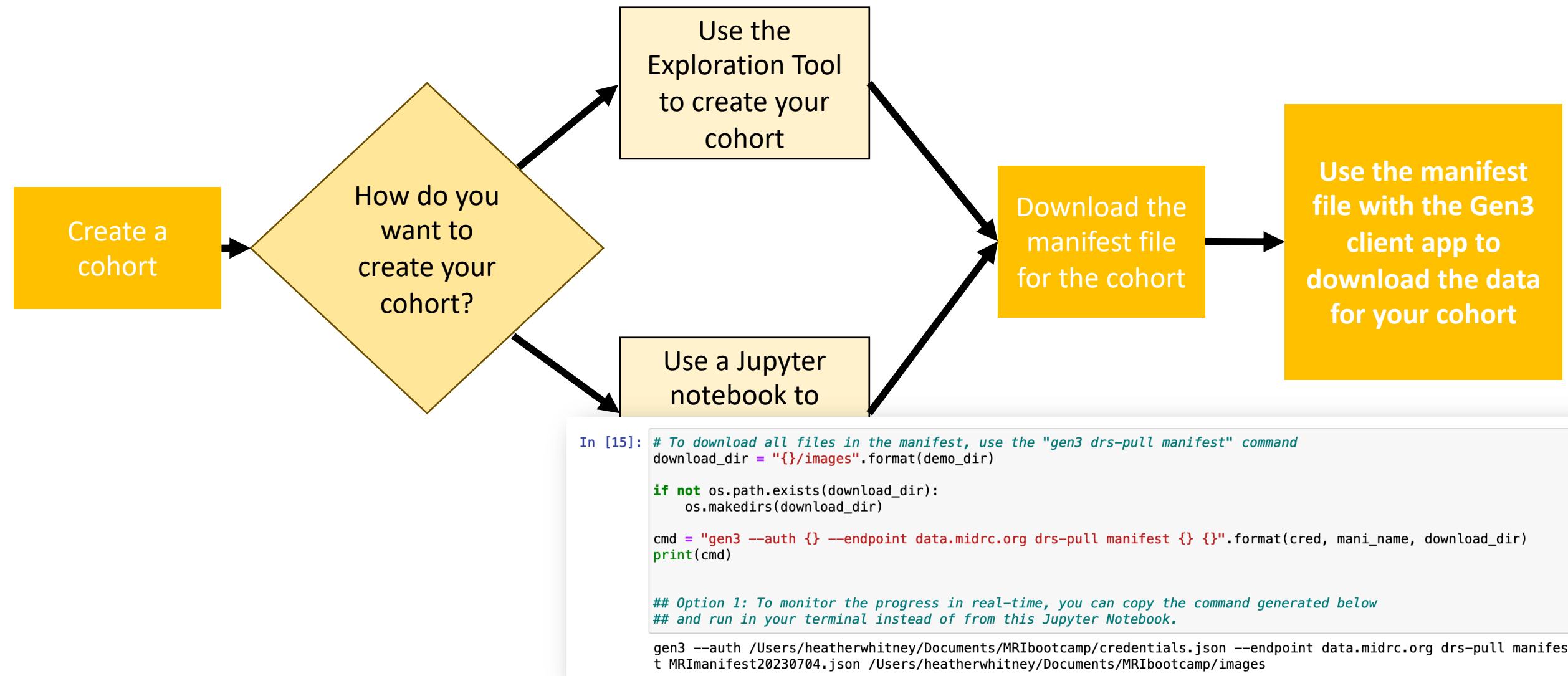
Developing a pipeline

Download



Developing a pipeline

Download



Developing a pipeline

Begin working with data