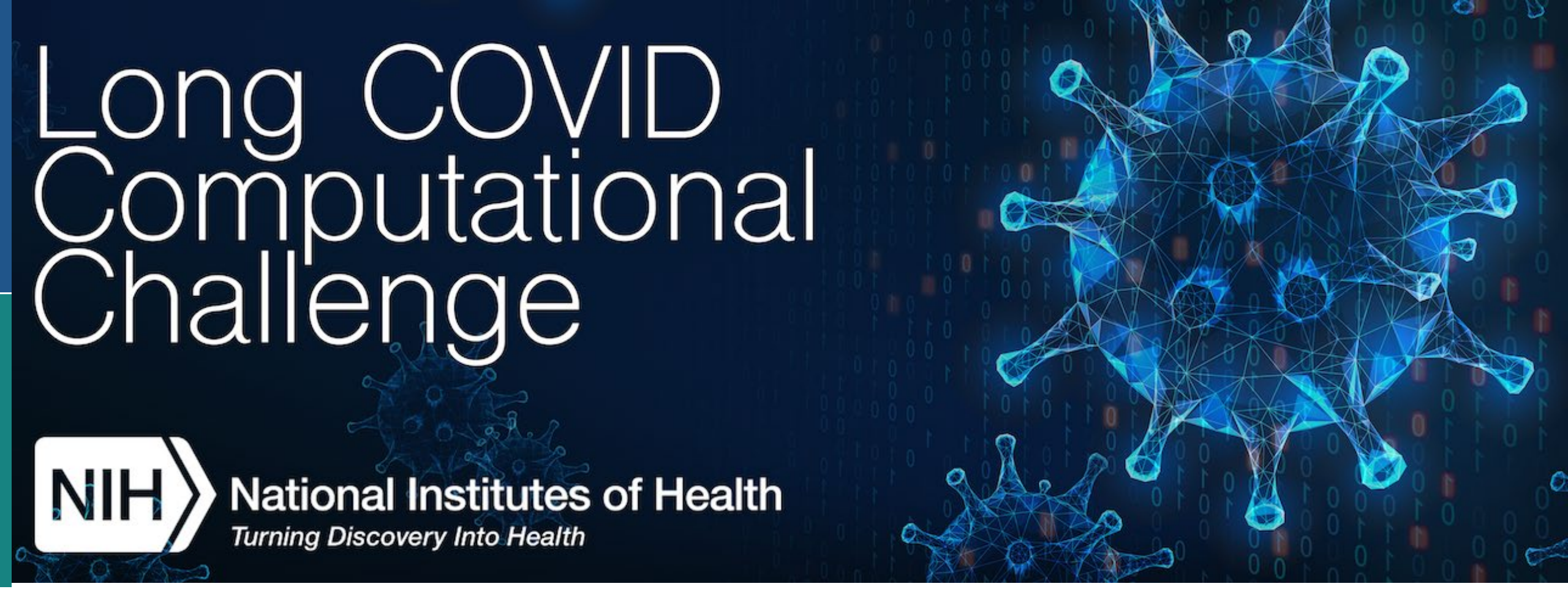


Long COVID Computational Challenge (L3C) Webinar

September 21, 2022

Zoom Rules of the Road

- This Webinar will be recorded
- All Participants, except the speakers and panelists, will be muted
- Auto Closed Captioning is available; to view, click the Live Transcript button
- Use the “Questions and Answers” function to ask questions and up-vote; questions will be answered during the Q&A session
- The recording and slides from today's webinar will be posted here: https://www.nih.gov/research-training/medical-research-initiatives/radx
- See Challenge.gov for more information on the L3C Challenge

Presenters and Panelists

Presenters:

Orlando Lopez, PhD (NIH/NIDCR)

Tim Bergquist, PhD (Sage Bionetwork)

Additional Panelists:

Josh Fessel, MD, PhD (NIH/NCATS)

Taylor Gilliland, PhD (NIH/NIBIB)

Erin Iturriaga, DNP, MSN, RN (NIH/NHLBI)

Asif Rizwan, PhD (NIH/NHLBI)

Ivonne Schulman, MD (NIH/NIDDK)

Acknowledgements

Special thanks to all those who contributed their time to this effort!

Erin Iturriaga	NHLBI	Bill Kapogiannis	NICHD
Ivonne Schulman	NIDDK	Susana Serrate-Sztein	NIAMS
Jue Chen	NHLBI	Asif Rizwan	NHLBI
Orlando Lopez	NIDCR	Taylor Gilliland	NIBIB
Amy Wernimont	CSR	Sweta Ladwa	NHLBI
Bonnie Joubert	NIEHS	Gabriel Anaya	NHLBI
Daniel Shaughnessy	NIEHS	Leonie Misquitta	NCATS
Danilo Tagle	NCATS	Rachel Scheinert	OD
Denny Buxton	NHLBI	Josh Fessel	NCATS
Joel Islam	OD	Timothy Bergquist	Sage Bionetwork
Qi Duan	NIBIB	Emily Pfaff	UNC
Yanli Wang	NLM	Johanna Loomba	UVA
Mary Pelleymounter	NINDS	Rick Woychik	NIEHS

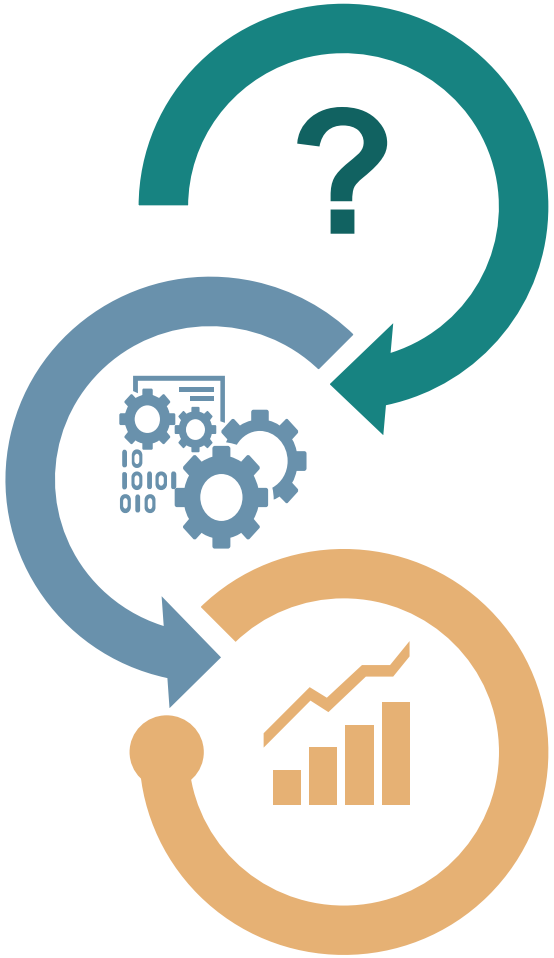
Bolded names indicate WG co-chairs

L3C Webinar Agenda

1. Background & Objectives 3:00 – 3:10 PM ET
2. Challenge Administration 3:10 – 3:30 PM ET
3. Q&A Session – 30 minutes 3:30 – 4:00 PM ET

Community Challenges

Format of a Challenge



The Question

Challenge organizers pose a research question

Build

Participants build methods to answer that question

Benchmark

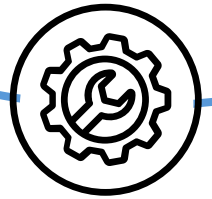
Challenge organizers benchmark those methods against a hidden gold standard

L3C Challenge Overview

PASC/Long COVID

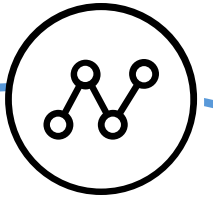
- Long term effects
- Persistent symptoms
- PASC*/ Long COVID can affect anyone

*PASC: Post-acute Sequalae of SARS-CoV-2 Infection



Long COVID Computational Challenge (L3C)

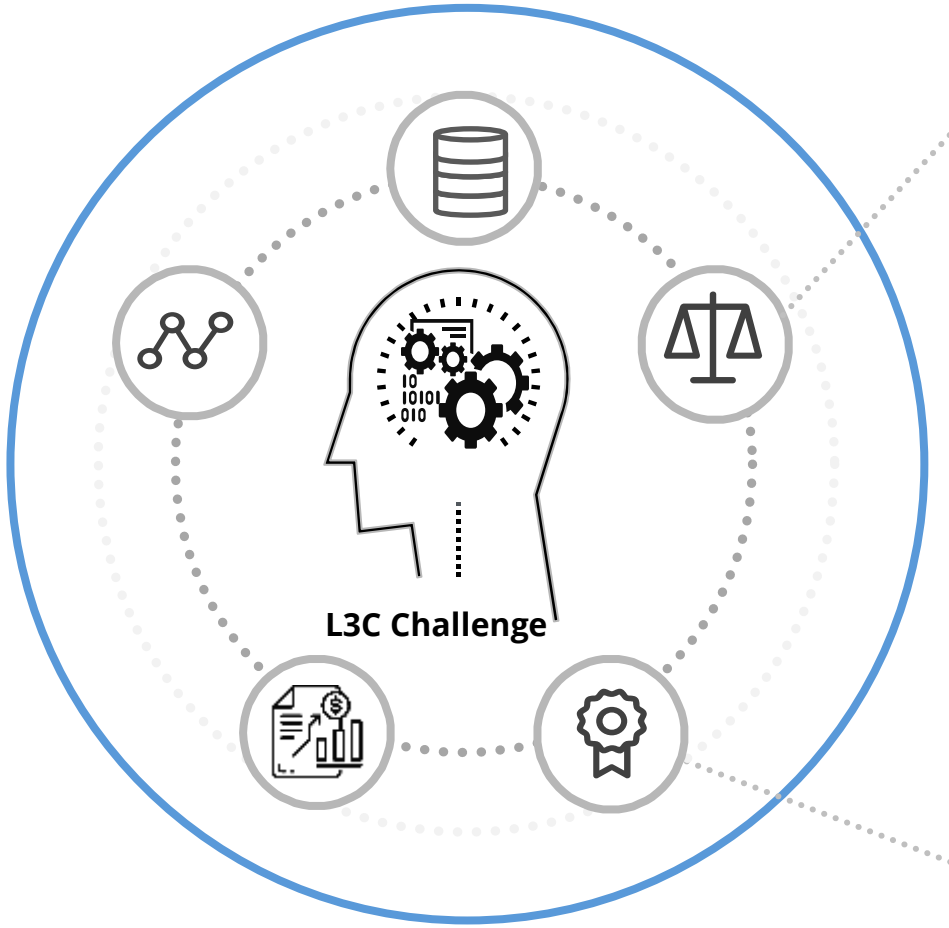
- Supports creative data-driven solutions
- Data-driven advancements to understand the risks of developing PASC/Long COVID



National COVID Cohort Collaborative (N3C)

- Leveraging N3C enclave
- AI-based prognostics for predicting risk of developing PASC/Long COVID

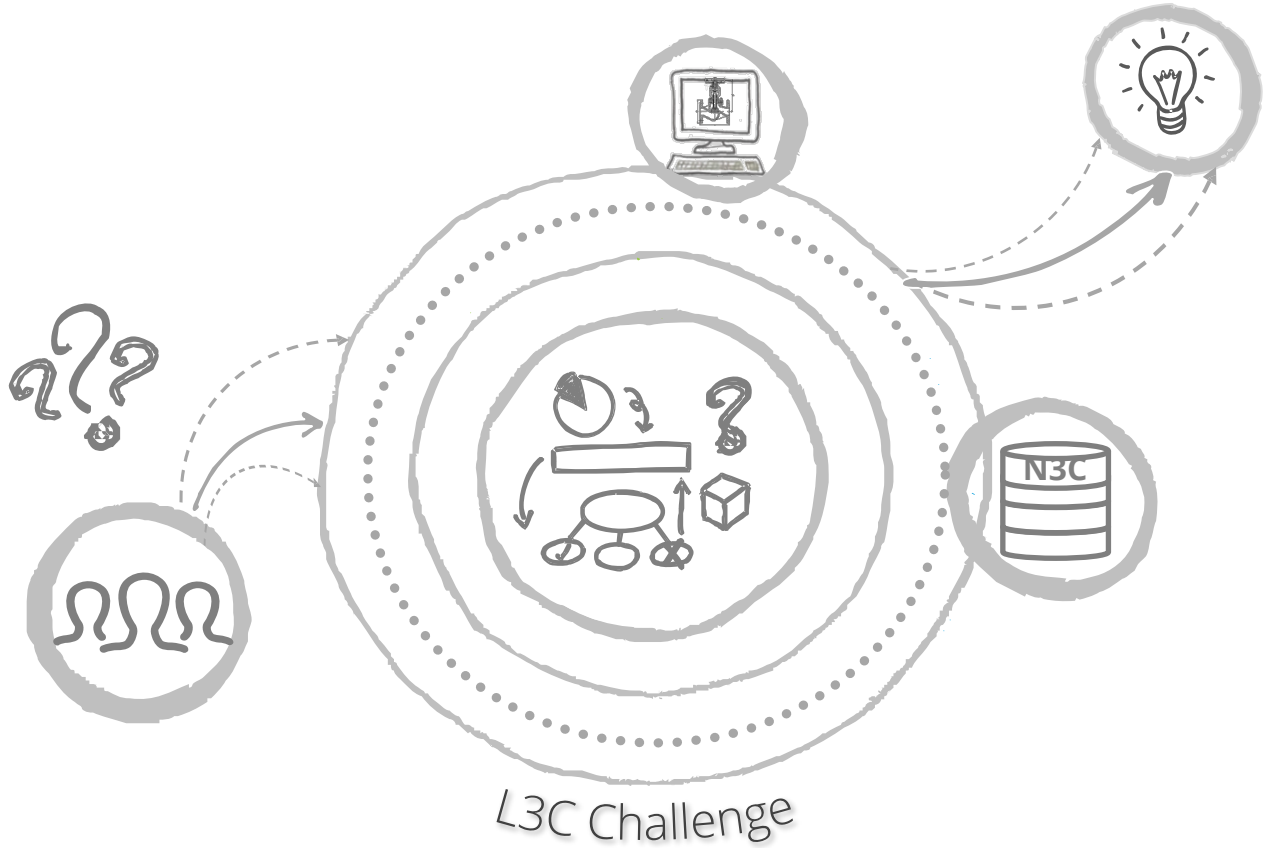
L3C Challenge Objective



- Identify patients likely to develop PASC/Long COVID
- Develop AI/ML model
- Open-source tools
- Using structured medical records
- Evaluate models using patients with ICD-10 code U09.9

L3C Challenge Question

If infected with SARS-CoV-2, who is more susceptible to developing PASC/Long COVID?



L3C Challenge Eligibility Rules

- Shall have **registered to participate in the Challenge under the rules** promulgated by the National Institutes of Health (NIH) as published in this announcement;
- Shall have **complied with all the requirements** set forth in this announcement;
- In the case of a private entity, **shall be incorporated in and maintain a primary place of business in the United States**, and **in the case of an individual**, whether participating singly or in a group**, shall be a citizen or permanent resident of the United States**. However, non-U.S. citizens and non-permanent residents can participate as a member of a team that otherwise satisfies the eligibility criteria**. Non-U.S. citizens and non-permanent residents are not eligible to win a monetary prize** (in whole or in part). Their participation as part of a winning team, if applicable, may be recognized when the results are announced.
- Shall **not be a federal entity or federal employee acting within the scope of their employment**;
- Shall **not be an employee of the Department of Health and Human Services** (HHS, or any other component of HHS) acting in their personal capacity;
- Who is **employed by a federal agency** or entity other than HHS (or any component of HHS), **should consult with an agency ethics official** to determine whether the federal ethics rules will limit or prohibit the acceptance of a prize under this Challenge;
- Shall **not be a judge of the Challenge, or any other party involved with the design, production, execution, or distribution** of the Challenge or the immediate family of such a party (i.e., spouse, parent, step-parent, child, or step-child).
- Shall be **18 years of age or older** at the time of submission.

Read Complete Rules - https://www.challenge.gov/?challenge=l3c&tab=rules
(Questions? Email RADxLongCOVIDChallengeAdmin@synapse.org)

L3C Challenge Participation Rules

Participation Rules of note

Open-Source Platform

- By participating in this Challenge, each Participant (whether an individual, group of individuals, or entity) selected to win a prize under this Challenge agrees to deposit their code for the models submitted to an open-source platform (e.g., GitHub)
- The code will be made available to the public under an open-source license
- Payment of the prize is contingent upon successful deposit of the code.
- As part of the submission requirement, each participant will submit their codes and a report summarizing their methods
- Sage will facilitate loading to open-source platform

N3C Enclave

- N3C Data Policies for Publication Governance Policies | N3C (cd2h.org)
- To access data within the N3C Data Enclave, an authorized institutional official from the Participant's home institution must sign the N3C Data Enclave Institutional Data Use Agreement (DUA) and agree to the terms of service outlined by NCATS and N3C
- Participants must agree and abide by N3C Data User Code of Conduct, Challenge Data Use Request, and complete required trainings.

Read Complete Rules - https://www.challenge.gov/?challenge=l3c&tab=rules
(Questions? Email RADxLongCOVIDChallengeAdmin@synapse.org)

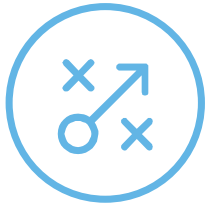
N3C Data Details

The data represents over 15 million patients

> 5.8 million COVID positive patients

> 17.5 billion rows of data

Data is de-identified to protect identifying patients



The dataset includes information such as

- Demographics
- Symptoms
- lab test results
- Procedures
- Medications
- Medical conditions
- Physical measurements

Observational Medical Outcomes Partnership (OMOP)

OMOP v 5.3 Common Data Model will be used for data schema and storage

https://ohdsi.github.io/CommonDataModel/cdm53.html

Diagnostic ICD-10 code U09.9 - Post COVID-19 Condition

U09.9 is a medical classification listed by WHO

Many patients in N3C are already identified by this code

Others may not have the code but could have undiagnosed PASC/Long COVID

N3C – L3C Challenge Data Access

Challenge Data

- Not all data available in the N3C enclave will be made available to participants in this challenge
- Participants will have access to all Long COVID patients available in a chosen data release version as well as randomly selected patients who have tested positive for COVID but have not been officially diagnosed with U09.9
- The data will represent a 1:4 ratio of Long COVID patients : non-Long COVID patients. This ratio may increase as the challenge goes on and teams develop more efficient workflows

Uncensored Training Data

All data that is available as of the chosen release version for the previously described challenge data

Censored Training Data

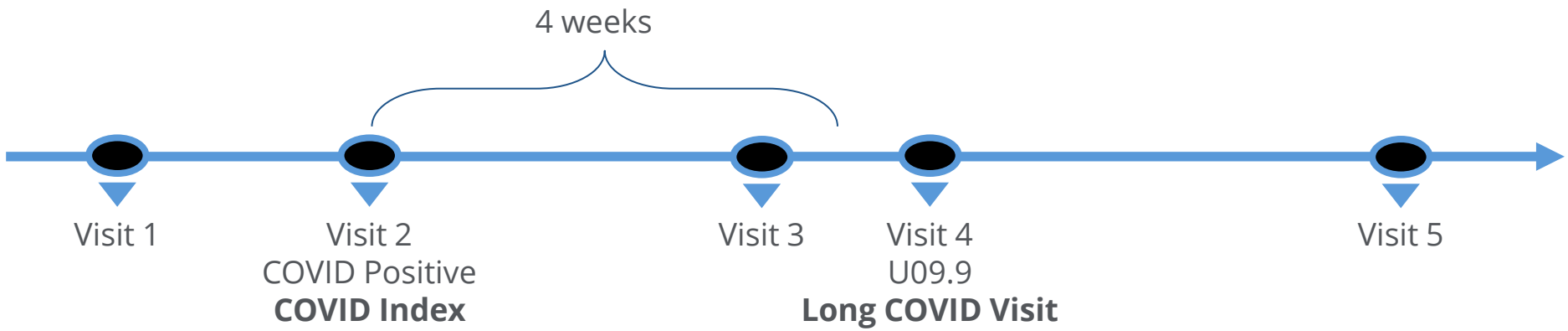
The challenge data will be censored, removing all data 4 weeks after the initial covid index date of the patients

Censored Testing Data

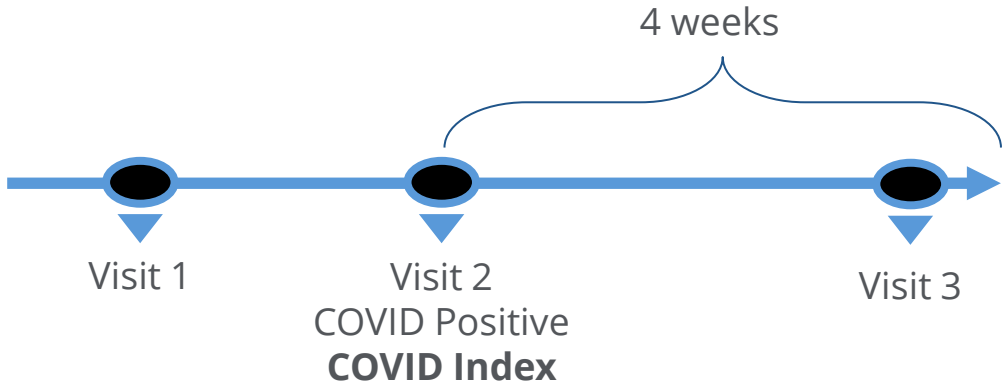
- The challenge data will be censored, removing all data 4 weeks after the initial covid index date of the patients
- This testing data will be a placeholder during the challenge and will be replaced with the new testing data once the model development phase is over

Long COVID Cohort Challenge Data

Uncensored Data

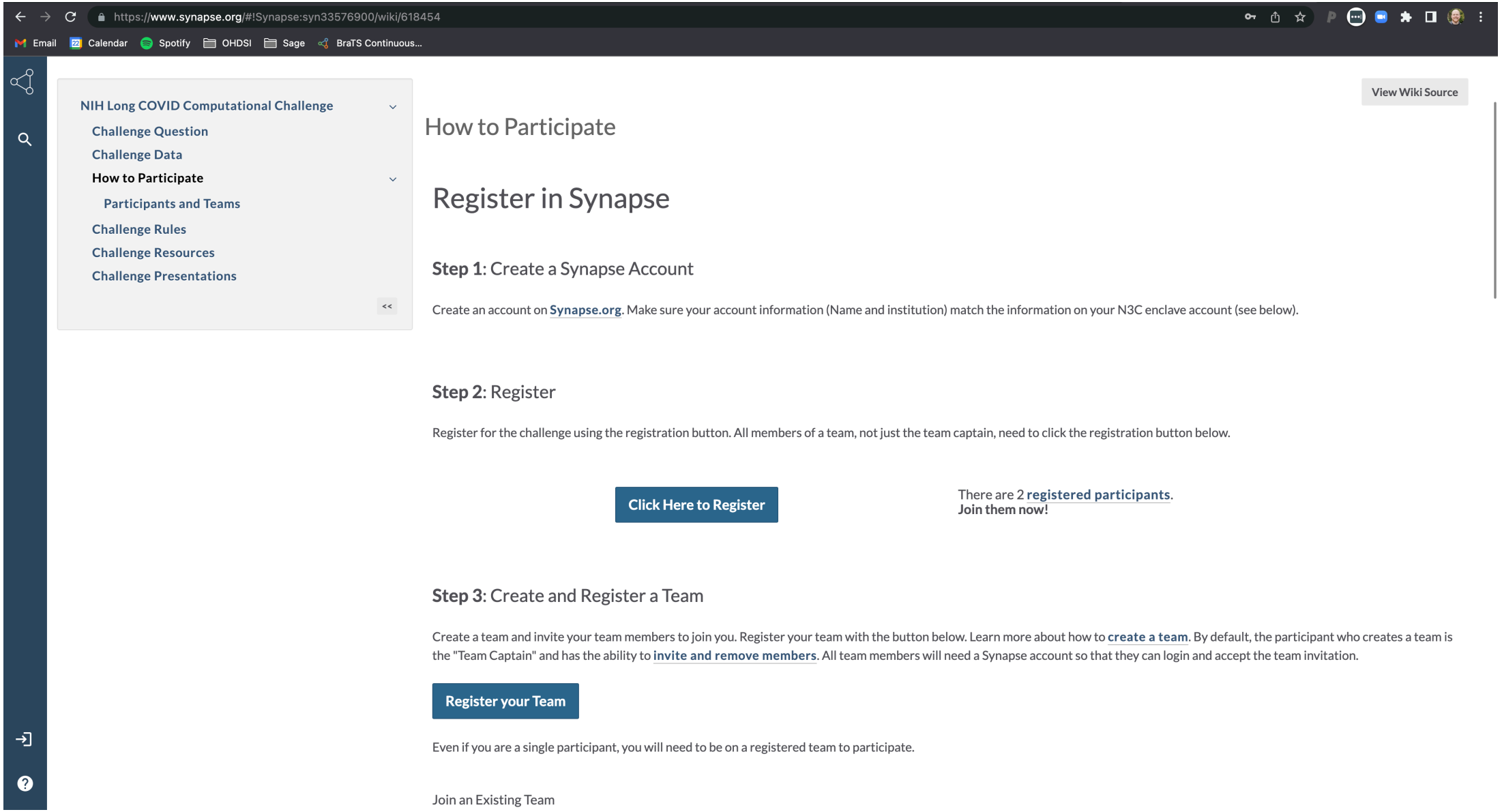


Censored Training/Testing Data



Challenge Registration

Register in Synapse



Challenge Registration

Register in N3C

https://www.synapse.org/L3C

May take **multiple weeks** to complete. DUAs and IRBs can take a while for institutions to process.

Has a much faster turnaround. About a **one week** turnaround.

Step 1: Signed DUA

In order to join this consortium, your organization will have to have a signed Data Use Agreement (DUA) with the N3C organization. Check if your institution has a DUA in place with the N3C organization. Citizen scientists will need to participate through an accessing institution.

Check DUAs

If not, have your institution sign a DUA with N3C.

Information for DUA

Citizen Scientists

If you are a citizen scientist and wish to participate in the challenge, you first should make every effort to become affiliated with an institution that can sign a DUA with N3C. If that fails, reach out to the **challenge administration team**, who will try to link you with a sponsoring institution who can sign a DUA. If that proves unsuccessful, the admin team will work with you to gain access to a synthetic dataset tailored to the challenge. If you are a citizen scientist and are not able to affiliate with an institution, you will have to follow the instructions on this website (**N3C Data Access Forms and Resources**).

Step 2: IRB Approval

Clarify with your institution if you need IRB approval to use de-identified clinical data for research per your institution's research policies.

Step 3: Register with the N3C organization.

Register with N3C

Step 4: Complete the required Training.

NIH Information Security Training

All researchers must complete the "**Information Security, Counterintelligence, Privacy Awareness, Records Management Refresher, Emergency Preparedness Refresher**" course (approx. 60-90 minutes), which can be accessed at NIH's information security training website. Be sure to **save evidence of completion for their records** (a screenshot or copy of the certificate of completion).

Complete NIH InfoSec Training

Human Subjects Research Training

Researchers must have completed their home institution's human subjects research training requirements. Researchers will be required to provide the date they completed training in their Data Use Request.

Step 5: Joining the Challenge Project

Request access to the **existing challenge project** by agreeing to the Terms of Use and submitting a Data Use Request for the challenge project space in the enclave. The project is titled `[RP-D5AE34] RADx Long COVID Prediction Challenge`. Participants will not be submitting a new Data Use Request and MUST use the existing DUR to request access.

Request Access

Possible Onboarding Timeline

Onboarding/Learning Curve Timeline

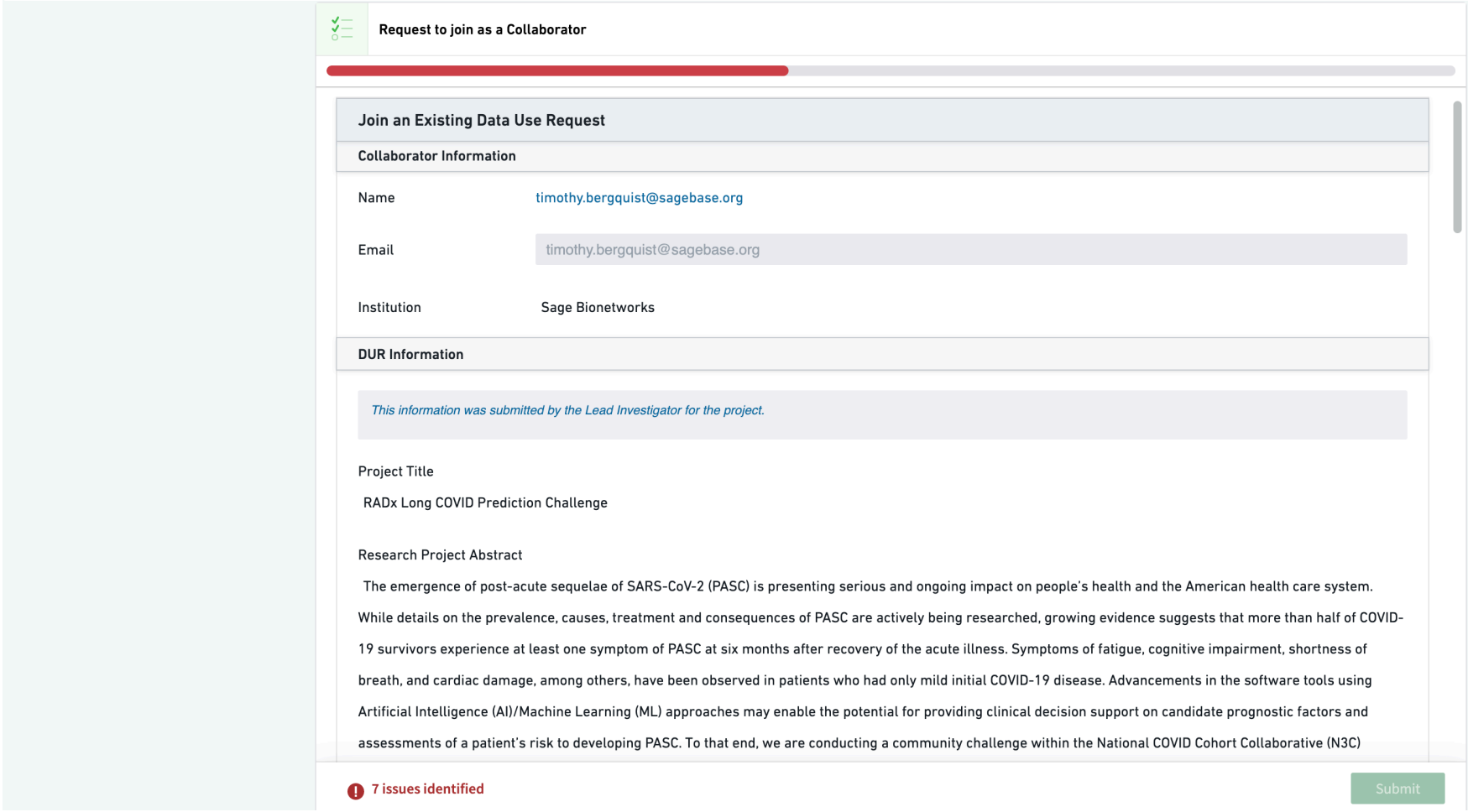
IRB and DUA paperwork submission	September 21, 2022
Completed N3C Onboarding	October 21, 2022
Tutorials/Learning Curve Completed	November 21, 2022
Judging Period Begins	December 16, 2022

Challenge Registration

Join the Existing DUR

National Covid Cohort Collective

Request to Join Existing DUR



Request to join as a Collaborator

Join an Existing Data Use Request

Collaborator Information

Name timothy.bergquist@sagebase.org

Email timothy.bergquist@sagebase.org

Institution Sage Bionetworks

DUR Information

This information was submitted by the Lead Investigator for the project.

Project Title

RADx Long COVID Prediction Challenge

Research Project Abstract

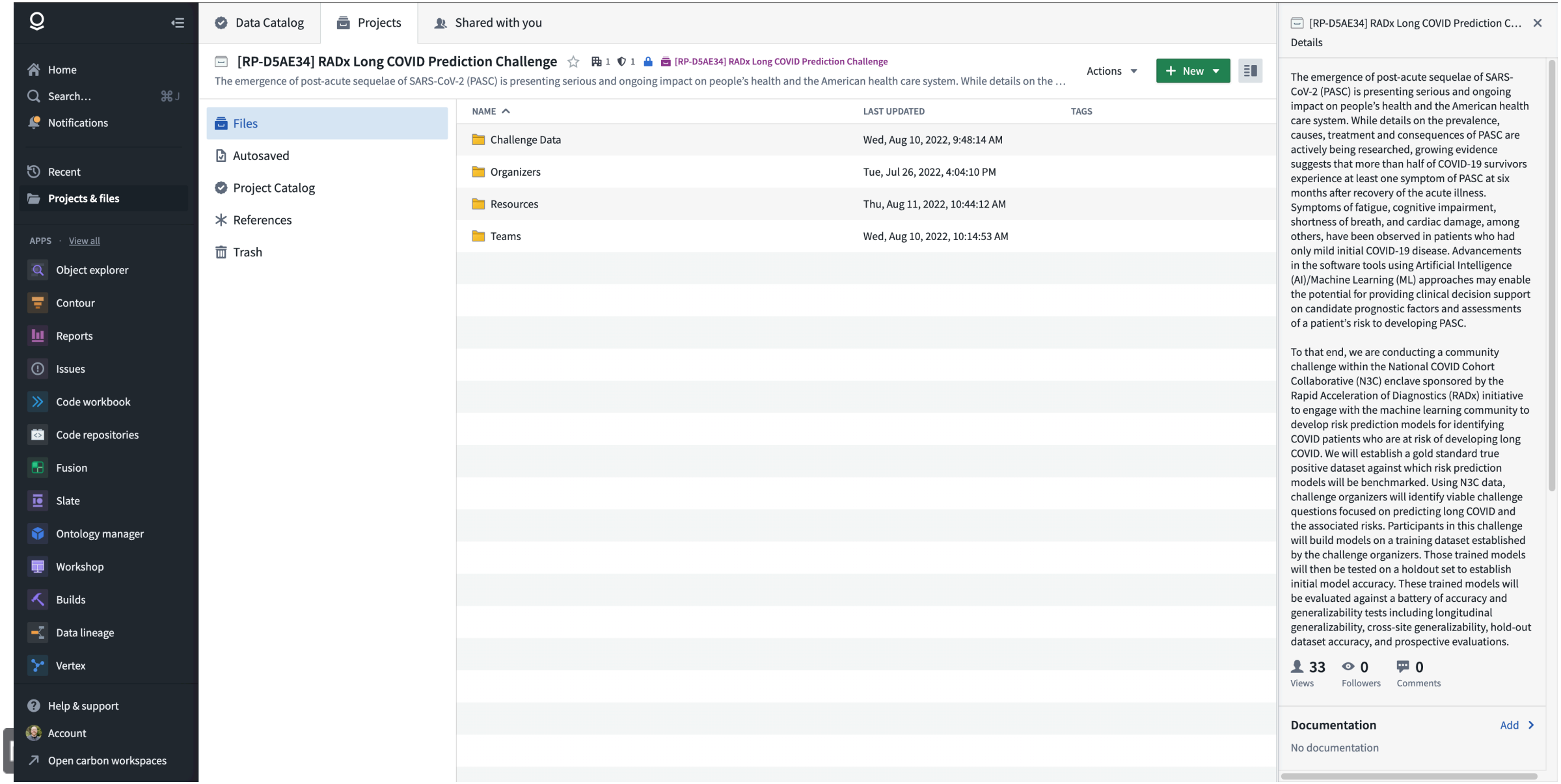
The emergence of post-acute sequelae of SARS-CoV-2 (PASC) is presenting serious and ongoing impact on people's health and the American health care system. While details on the prevalence, causes, treatment and consequences of PASC are actively being researched, growing evidence suggests that more than half of COVID-19 survivors experience at least one symptom of PASC at six months after recovery of the acute illness. Symptoms of fatigue, cognitive impairment, shortness of breath, and cardiac damage, among others, have been observed in patients who had only mild initial COVID-19 disease. Advancements in the software tools using Artificial Intelligence (AI)/Machine Learning (ML) approaches may enable the potential for providing clinical decision support on candidate prognostic factors and assessments of a patient's risk to developing PASC. To that end, we are conducting a community challenge within the National COVID Cohort Collaborative (N3C)

7 issues identified

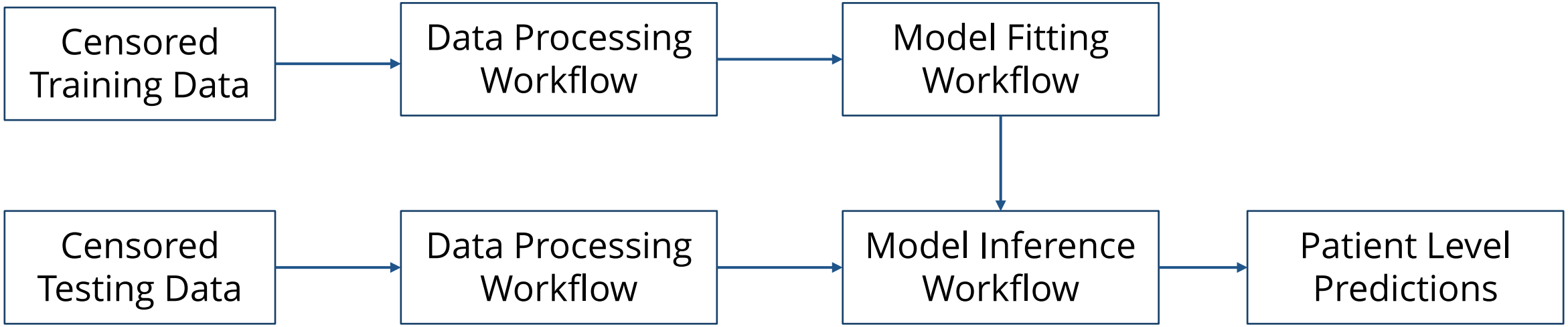
Submit

https://unite.nih.gov/workspace/slate/documents/collaborator-data-use-request?DUR=DUR-E6CBB51

Project Space



Challenge Submission



Example Code is available in the “Resources” folder.

We will be going through this workflow in an upcoming orientation.

Evaluation Metrics

An example of the metrics used to evaluate the models will include…

Quantitative Metrics

- Calibration metrics (Mean error, Brier score, etc.)
- Area Under the Precision Recall Curve
- Area Under the Receiver Operator Curve
- Precision and Recall

Qualitative Metrics

Utility

- Feature Interpretability and Relevance
- Timeliness of Predictions
- Context Utility
- Likelihood of Implementation

Reproducibility

- Technical Reproducibility
- Prediction Reproducibility
- Documentation Reproducibility
- Method Clarity

Prizes

The total prize purse for this challenge is up to $500,000

$200,000

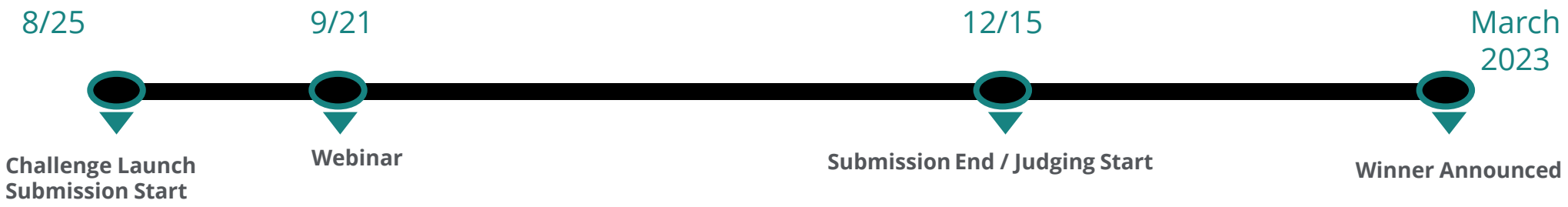
$150,000

$100,000

$50,000 **Honorable Mention Prize Pool**

Up to **5** projects may be selected for an **honorable mention** prize at the sole discretion of the challenge managers. Honorable mention prize winners will receive an **equal share** of the Honorable Mention prize pool.

Challenge Timeline



Challenge Launch	August 25, 2022
Challenge Webinar	September 21, 2022, 3:00 PM Eastern Time
Submission Period	August 25 - December 15, 2022
Judging Period	December 16, 2022 – February 15, 2023
Winner Announced	March 2023

Additional Learning Opportunities

Tutorials and Orientations

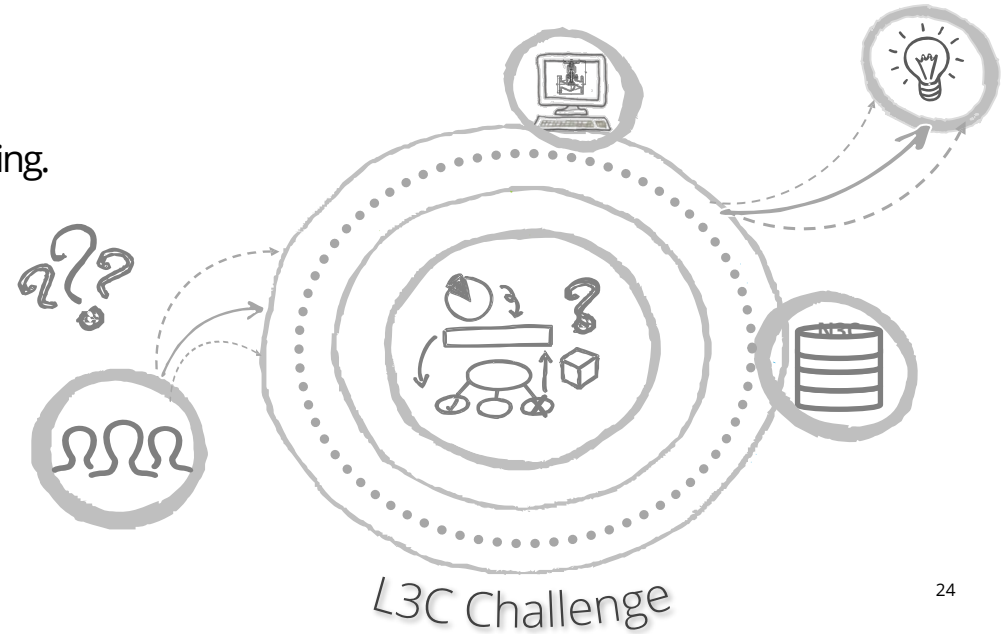
- Over the next few weeks, we will be hosting tutorials detailing techniques and methods for building efficient models, for working with the N3C data, for understanding the caveats of the N3C data, and other topics.
- Details on time and dates will be announced soon. These sessions will also be recorded and made available.

Resources

There are additional documents in the **Resources** folder of the enclave project that have *links to tutorials*, *challenge instructions*, and *submission examples* that you should look through.

Office Hours

- We will also be hosting regular office hours if you'd like to swing by for a more informal setting.
- **Thursdays at 10am PST, starting on September 29th,**
- https://us02web.zoom.us/meeting/register/tZIsceqtrjoqHNTfFgPUP0etb4saRxEQqlMY



Questions & Answers