Collaborative
Cohort of Cohorts
for COVID-19 Research

C4R Analysis Commons Training

August 11, 2021

Pallavi Balte, MBBS, PhD

Associate Research Scientist
Department of Medicine| Division of General Medicine
Columbia University Irving Medical Center, NY

Agenda

- Overview of C4R and BioData Catalyst/Seven Bridges (3:00 – 3:30pm)
  - Data availability on C4R Analysis Commons
  - Data harmonization in C4R
  - BioData Catalyst/Seven Bridges
    - How to get an account?
    - Organization and access to C4R Analysis Commons
    - Data security
- Live Demo in BioData Catalyst/Seven Bridges (3:30 – 4:00pm)
- Q and A session (4:00 – 5:00pm)

Target population (N=14 cohort studies)

Original cohort focus

Heart

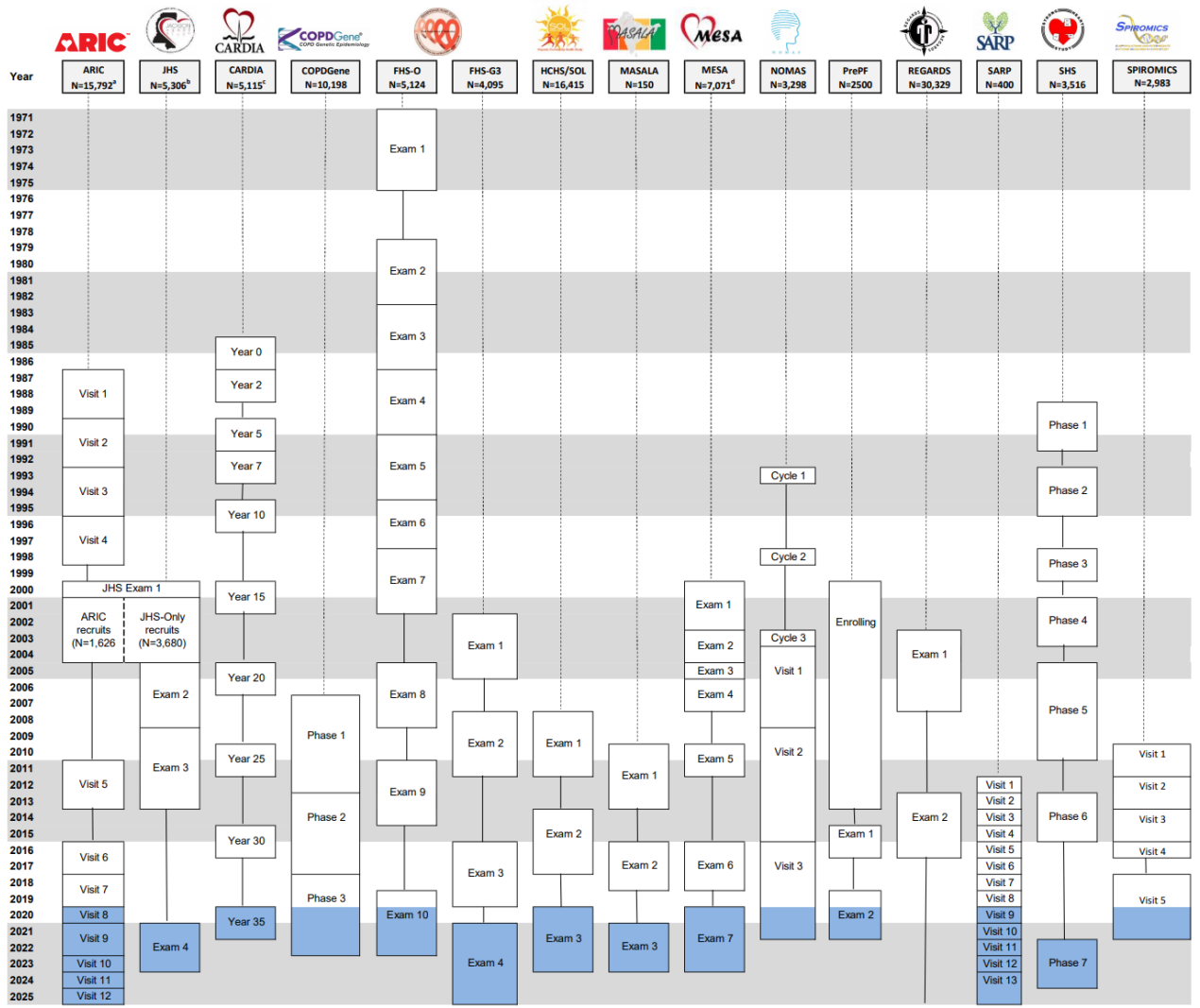
Lung

Brain

Cohort	N
ARIC	5046
CARDIA	4221
COPDGene	4000
FIP	2500
Framingham	7258
HCHS-SOL	8400
JHS	2317
MASALA	500
MESA	4683
NOMAS	1267
REGARDS	8000
SARP	380
SPIROMICS	1800
SHS	2701
Total	53073

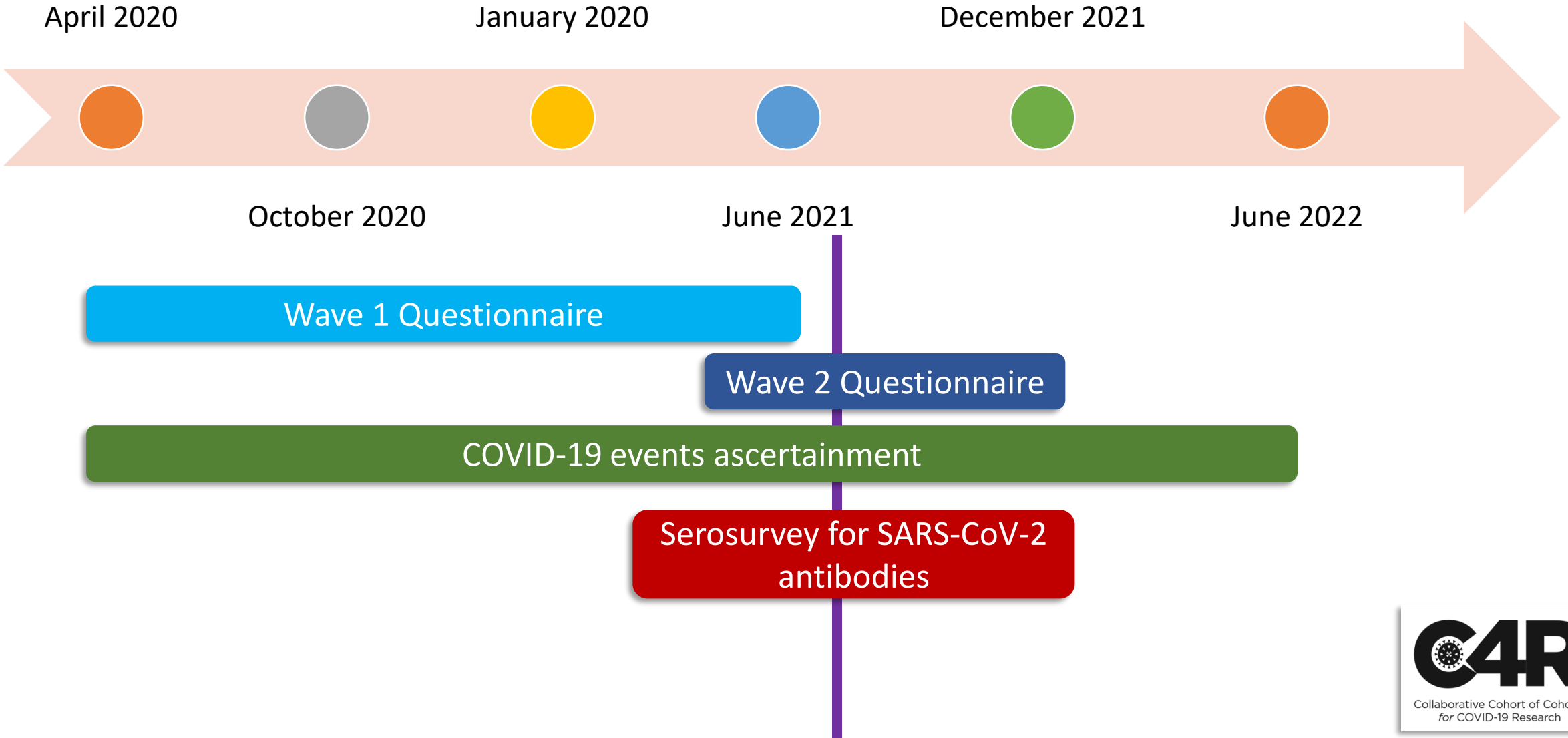
Over 1,000,000 person-years of follow-up

- Repeated measures across multiple organ systems, many biomarkers, and ‘Omics
- Pandemic era exams already funded in numerous C4R cohorts



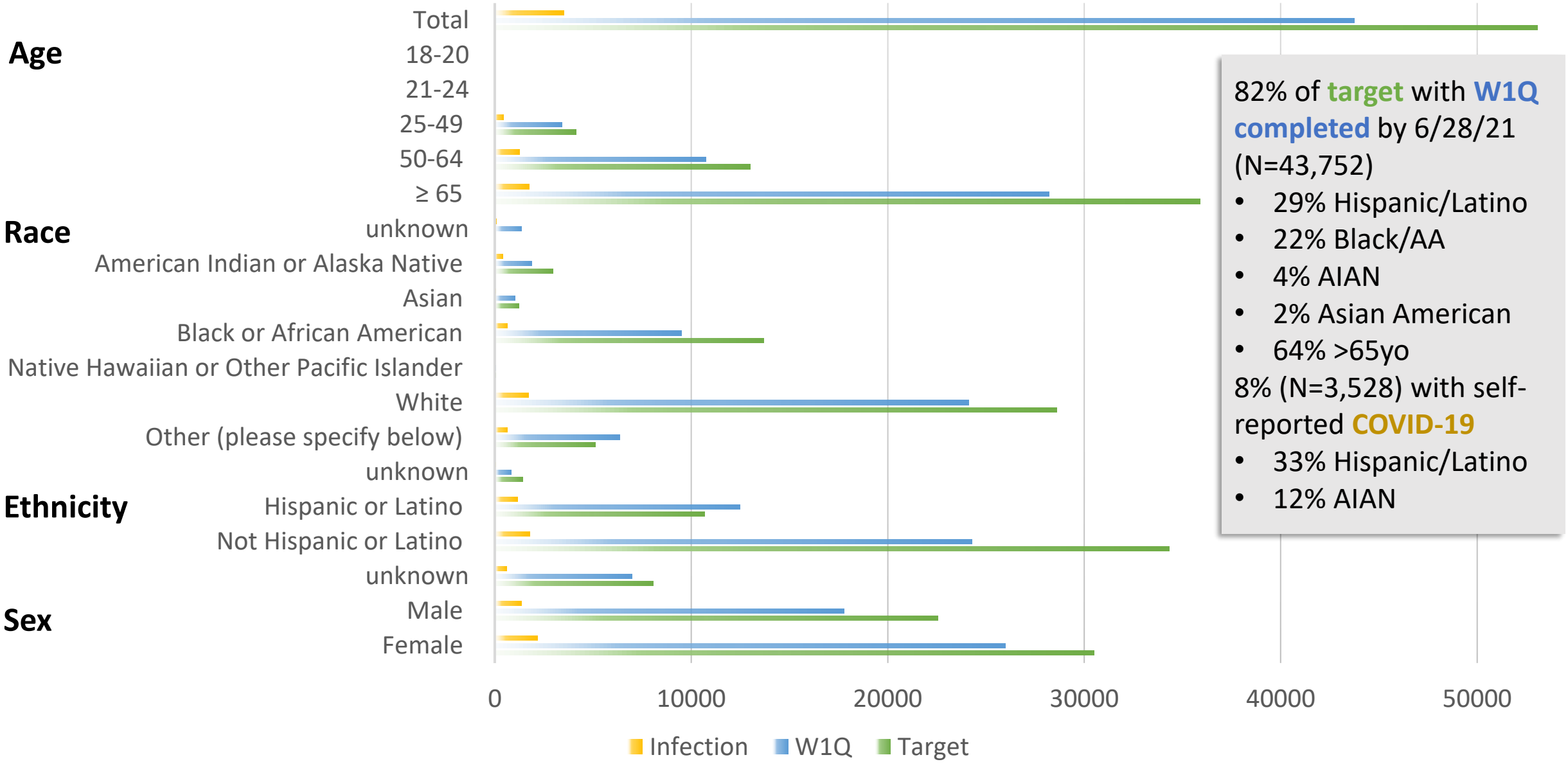
Oelsner EC et al, https://www.medrxiv.org/content/10.1101/2021.03.19.21253986v1

C4R Study Timeline



Collaborative Cohort of Cohorts *for* COVID-19 Research

Wave 1 questionnaire completion (June 28, 2021)



82% of **target** with **W1Q completed** by 6/28/21 (N=43,752)

- 29% Hispanic/Latino
- 22% Black/AA
- 4% AIAN
- 2% Asian American
- 64% >65yo

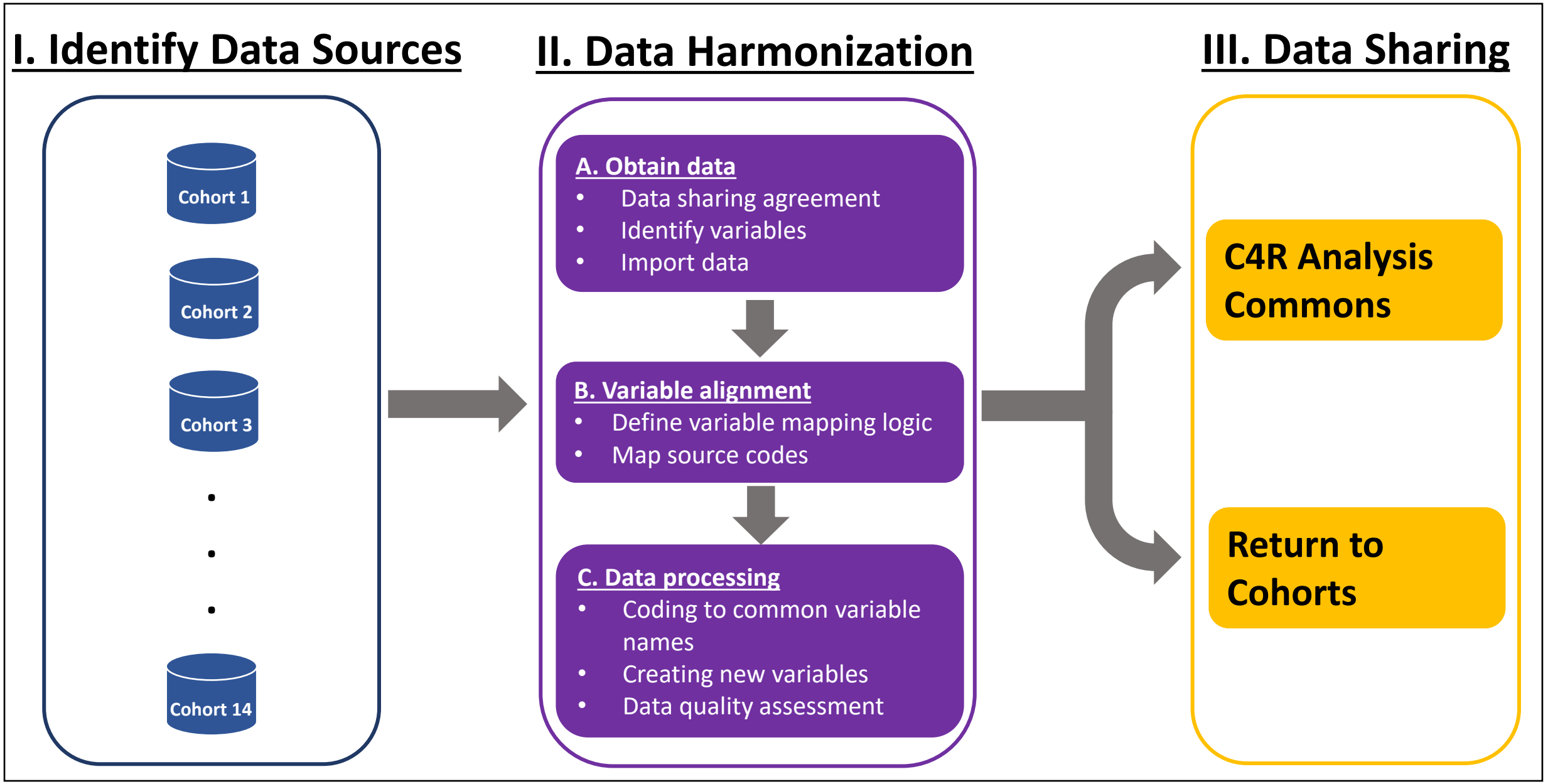
8% (N=3,528) with self-reported **COVID-19**

- 33% Hispanic/Latino
- 12% AIAN

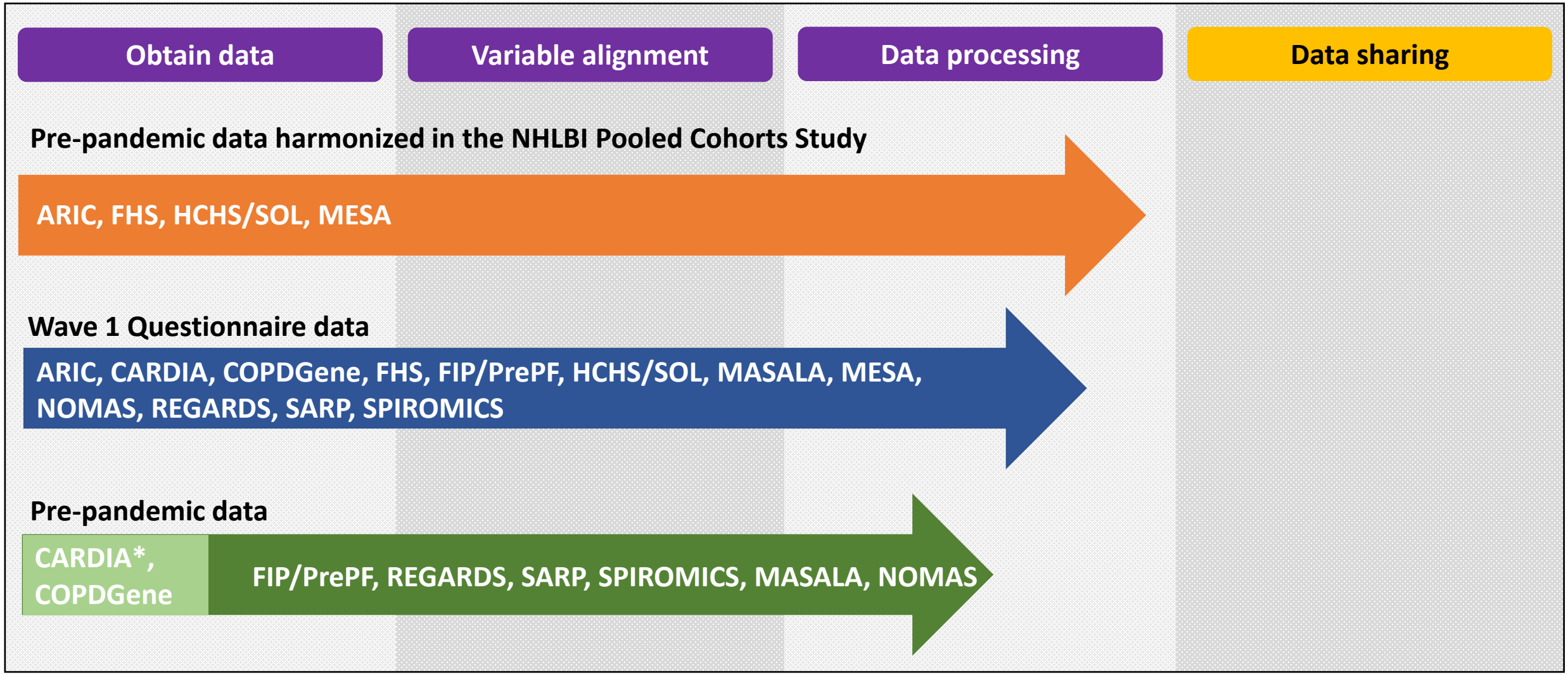
C4R Data

- COVID-19 Endpoints
  - Wave 1 Questionnaire (W1Q)
- Dried Blood Spots (DBS) data
- Adjudicated COVID-19 events
- Socio-demographics
  - Most recent pre-pandemic visit
  - Longitudinal data
- Pre-pandemic measures
  - Most recent pre-pandemic visit
  - Longitudinal data
- Adjudicated cardiac and pulmonary events

Data harmonization within C4R Analysis Commons



Harmonization tracking in C4R Analysis Commons



DUAs are pending with JHS and SHS

W1Q COVID-19 Endpoints Variables

	ARIC	CARDIA	COPDGene	FHS	FIP/PrePF	HCHS/SOL	MASALA	MESA	NOMAS	REGARDS	SARP	SPIROMICS
COVID_Infection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
COVID_Selfreport	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
COVID_HCP	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
COVID_Hospitalized	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
COVID_Testpositive	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
COVID_severity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
COVID_reinfection				✓	✓		✓		✓			
COVID_symptoms	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
COVID_ICU				✓	✓		✓	✓	✓			✓
COVID_recover	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
COVID_vaccine	✓			✓	✓		✓		✓			

DUAs are pending with JHS and SHS

Dried Blood Spot Variables

- Antibodies against the spike (S) or nucleocapsid (N) proteins of the SARS-CoV-2 virus
- Natural COVID-19 infection → *Antibodies against both N and S proteins*
- COVID-19 vaccination → *Antibodies against only S protein*

<u>Threshold for Reactive test</u>:
Antibodies against N protein: ≥ 5000 AU/mL
Antibodies against S protein : ≥ 1000 AU/mL

C4R variable	Explanation
Test_result	Reactive, Non-reactive, Indeterminate, Test cancelled
Num_blood_spots	Number of DBS: 1-5
MFI_N	IgG antibody titers against N protein
MFI_S	IgG antibody titers against S protein
MFI_N_reactive	Reactive to N protein based on threshold of 5000Au/mL: yes, no
MFI_S_reactive	Reactive to S protein based on threshold of 1000Au/mL: yes, no
Test_date	Date of blood spots collection

MFI: Mean fluorescence intensity units

Adjudicated COVID-19 Events Variables

- COVID-related hospitalizations and deaths
  - ICD codes
  - Medical records
- Level of certainty for each diagnosis
  - Definite
  - Probable
  - Not applicable
- Additional data
  - Vital signs (RR, O2, O2 supplementation)
  - Medications

Diagnosis	Severity	Complications
COVID-19…	Infection	Pneumonia
	Hospitalization	Myocardial infarction
	Severe illness	Stroke
	Critical illness	Pulmonary embolism
	Fatal illness	DVT
		Renal failure

Core pre-pandemic measures

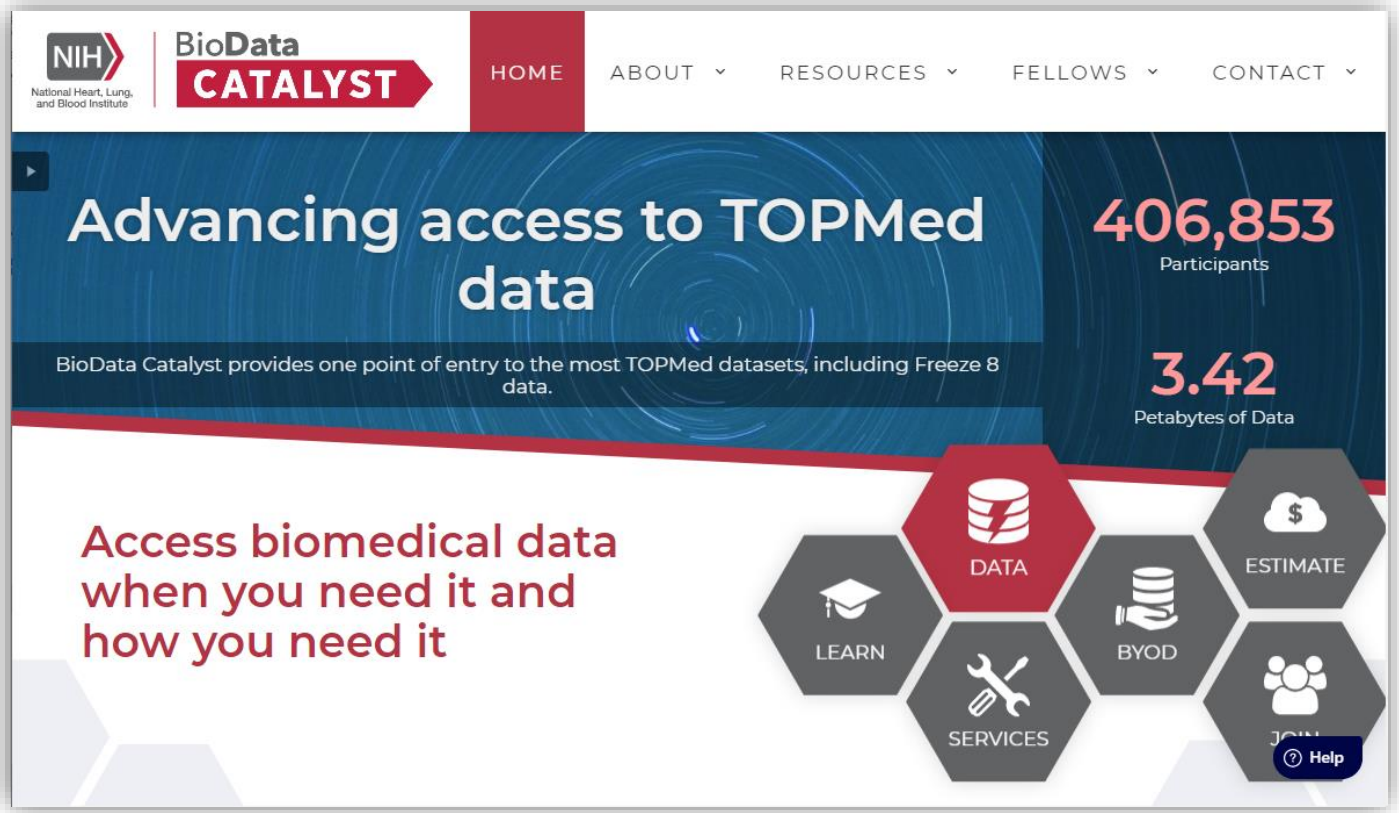
- *Socio-demographics*: age, sex, race/ethnicity, education, birth year
- *Anthropometry*: height, weight, BMI, hip and waist circumference
- *Smoking history*: smoking status, pack-years, cigarettes per day
- *Spirometry*: pre- and post-bronchodilator measurements, QC variables
- *Past medical history*: hypertension, diabetes, CVD, CKD, cancer, etc.
- *Medications*: anti-hypertensives, insulin, oral hypoglycemics, statins, steroids, aspirin, etc.
- *Blood pressure measurements*: systolic and diastolic
- *Labs*:
  - Renal biomarkers: serum creatinine, uACR, eGFR
  - Lipids: HDL, LDL, total cholesterol, triglycerides
  - Inflammatory biomarkers: CRP, fibrinogen, D-dimer
  - Blood glucose levels, HbA1c
- *Respiratory symptoms*: cough, phlegm, shortness of breath, etc.
- *Adjudicated cardiac events*: MI, CVD, CHF, Stroke, etc.
- *Adjudicated pulmonary events*: asthma, emphysema, chronic bronchitis, COPD

Socio-demographic Characteristics of C4R population

	W1Q completion		Infections to date	
	N	%	N	%
Total	43,752	100%	3,528	100%
Age				
18-20	17	0%	3	0%
21-24	14	0%	0	0%
25-49	3,415	8%	439	12%
50-64	10,727	25%	1,261	36%
≥ 65	28,210	64%	1,743	49%
unknown	1,369	3%	82	2%
Race				
American Indian or Alaska Native	1,881	4%	411	12%
Asian	1,036	2%	34	1%
Black or African American	9,501	22%	647	18%
Native Hawaiian or Other Pacific Islander	34	0%	5	0%
White	24,125	55%	1,711	48%
Other (please specify below)	6,356	15%	653	19%
unknown	819	2%	67	2%
Ethnicity				
Hispanic or Latino	12,474	29%	1,151	33%
Not Hispanic or Latino	24,283	56%	1,770	50%
unknown	6,995	16%	607	17%
Sex Assigned at Birth				
Male	17,770	41%	1,360	39%
Female	25,979	59%	2167	61%

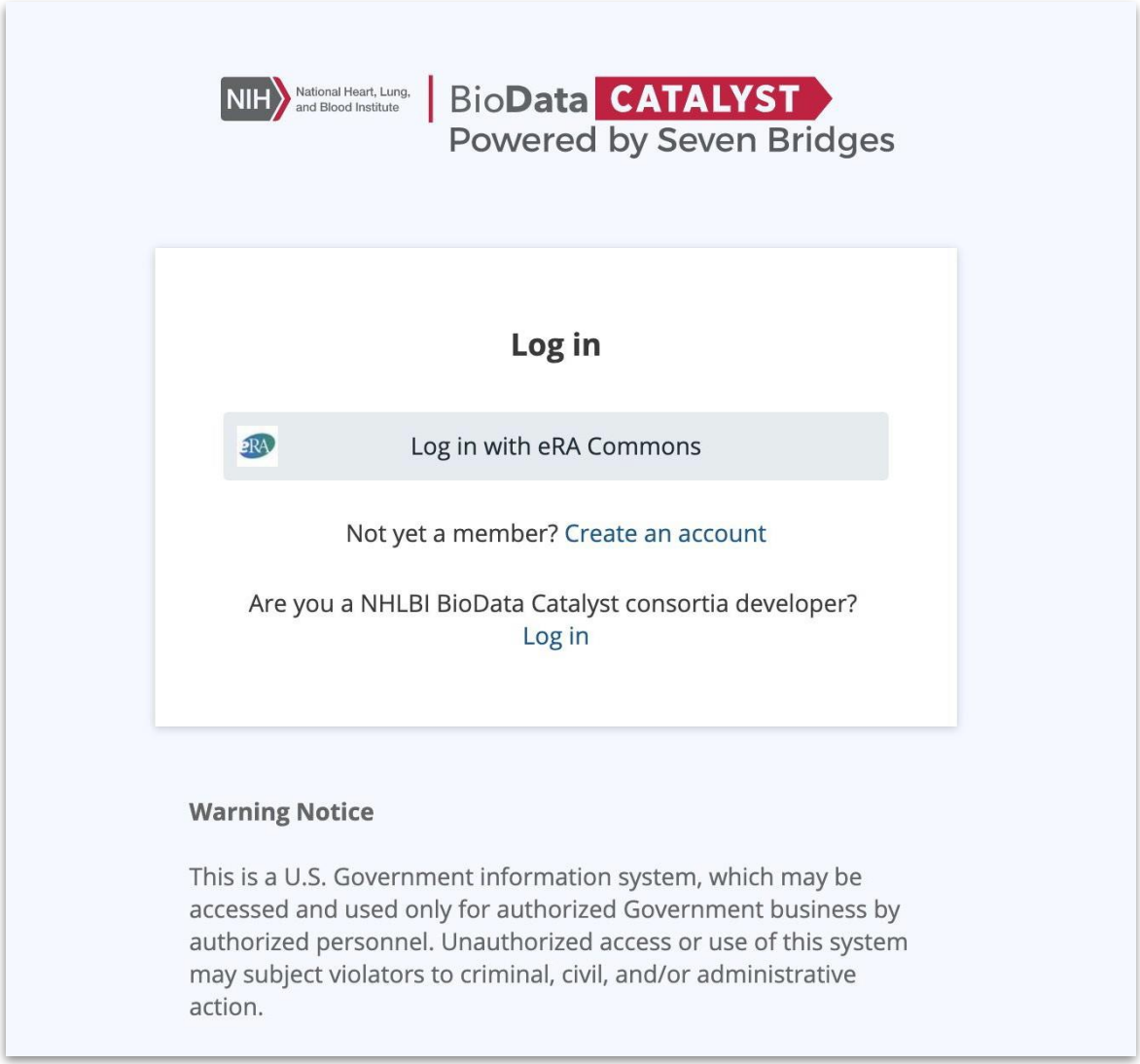
C4R Analysis Commons:
BioData Catalyst powered by Seven Bridges

- Cloud-based data sharing and analysis platform
- Secure place to upload data and share with the C4R DCHC as well as researchers/investigators at other institutes
- Analysis software: SAS Studio, R Studio, JupyterLab
- https://biodatacatalyst.nhlbi.nih.gov/
- Who can sign up?

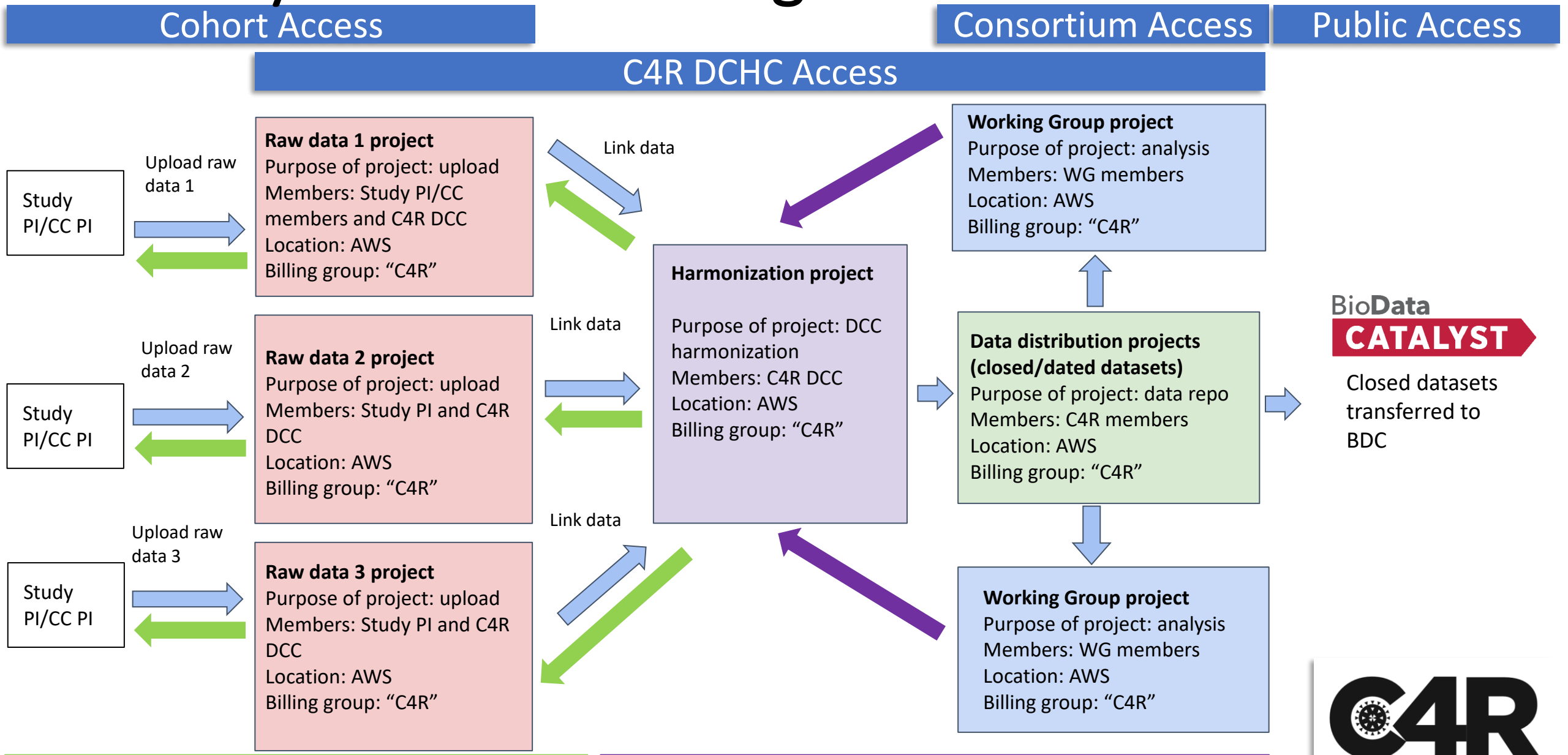


How to get an account on BioData catalyst/Seven Bridges?

- https://platform.sb.biodatacatalyst.nhlbi.nih.gov/
- Connect with your eRA Commons ID
- Getting Started Guide
- C4R DCHC will add you to the appropriate projects with the C4R billing group
- You can test out the platform using your $500 of pilot funding



C4R Analysis Commons: Organization and Access



Cohorts receive their own harmonized or derived C4R variables, not those from other cohorts

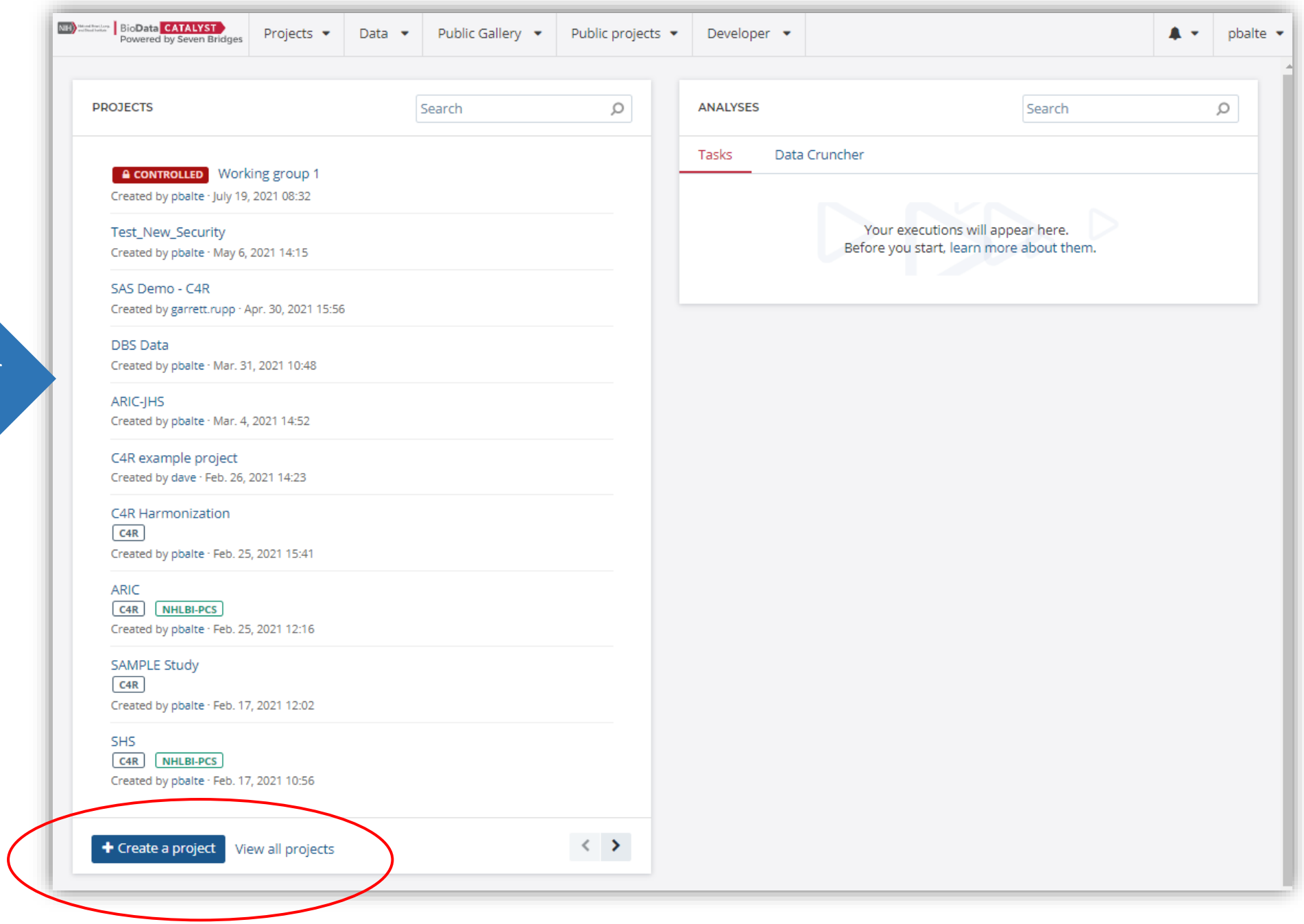
Working groups harmonize necessary additional variables and share back with main project for inclusion in future datasets

C4R
Collaborative Cohort of Cohorts *for* COVID-19 Research

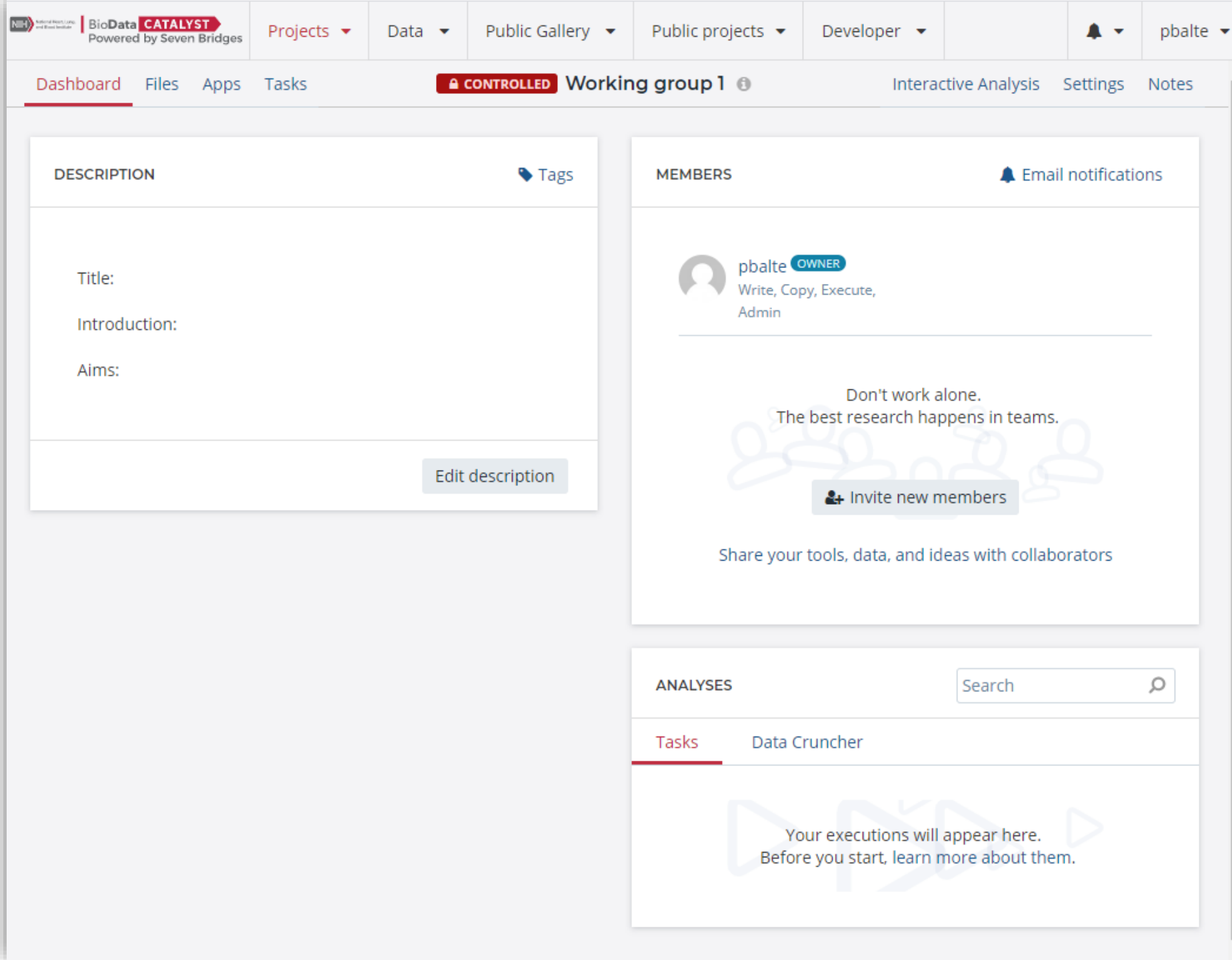
BioData Catalyst/Seven Bridges: Data Security

- Users are prohibited from downloading any controlled access, individual-level data
- Project privileges
  - Write
  - Copy
  - Execute
  - Admin
- BioData Catalyst Security Statement

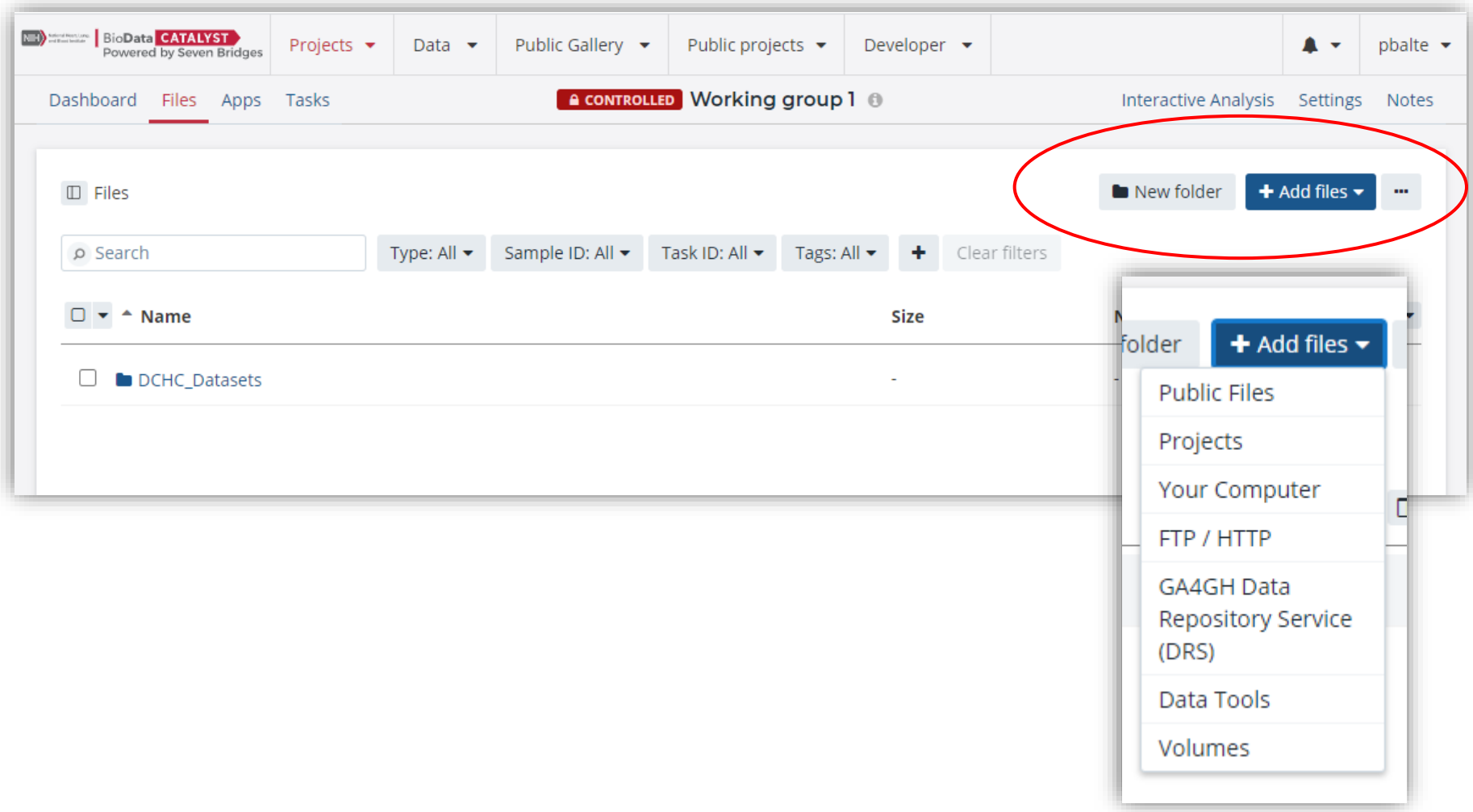
How to create a project?



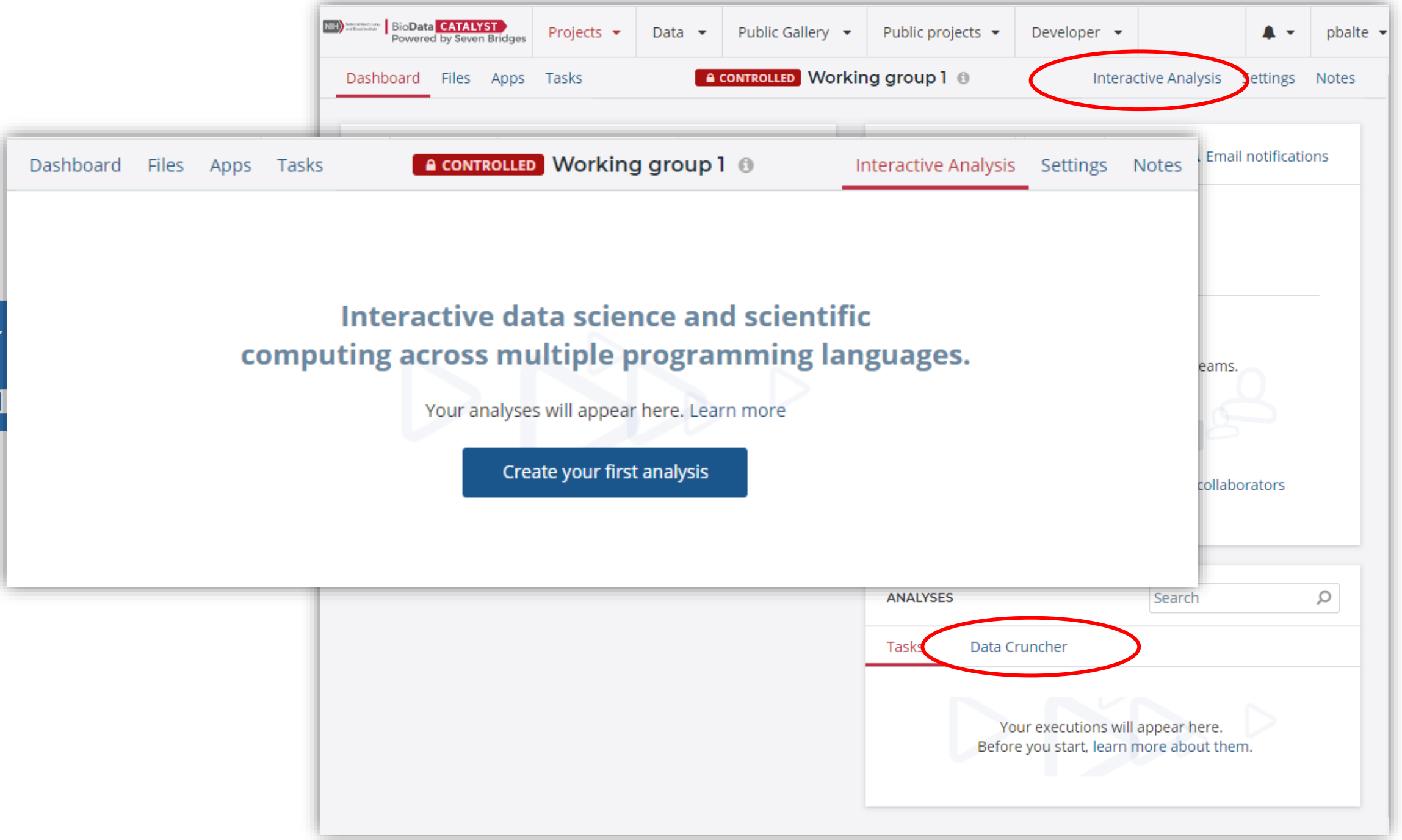
Project Dashboard



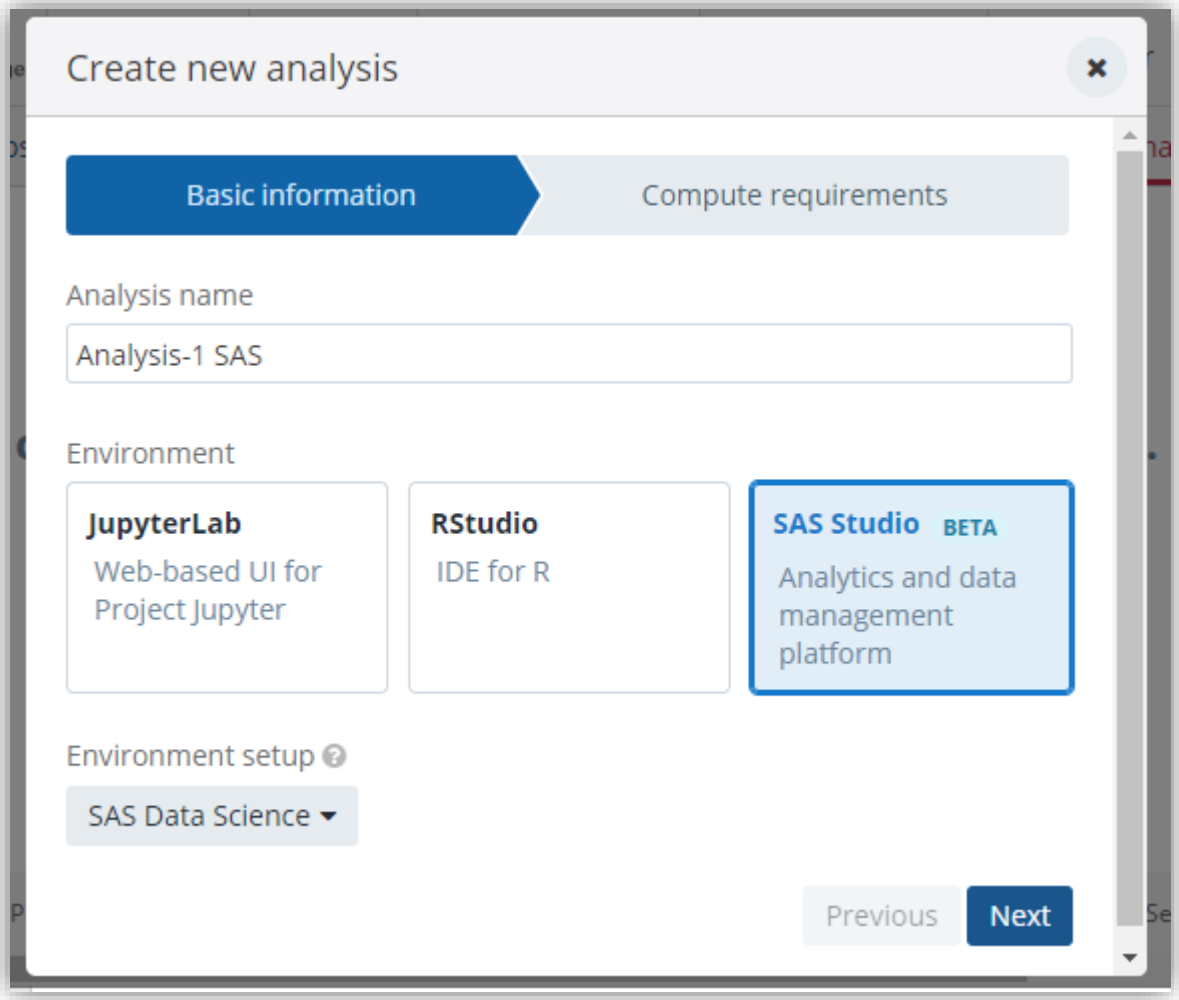
Project Files

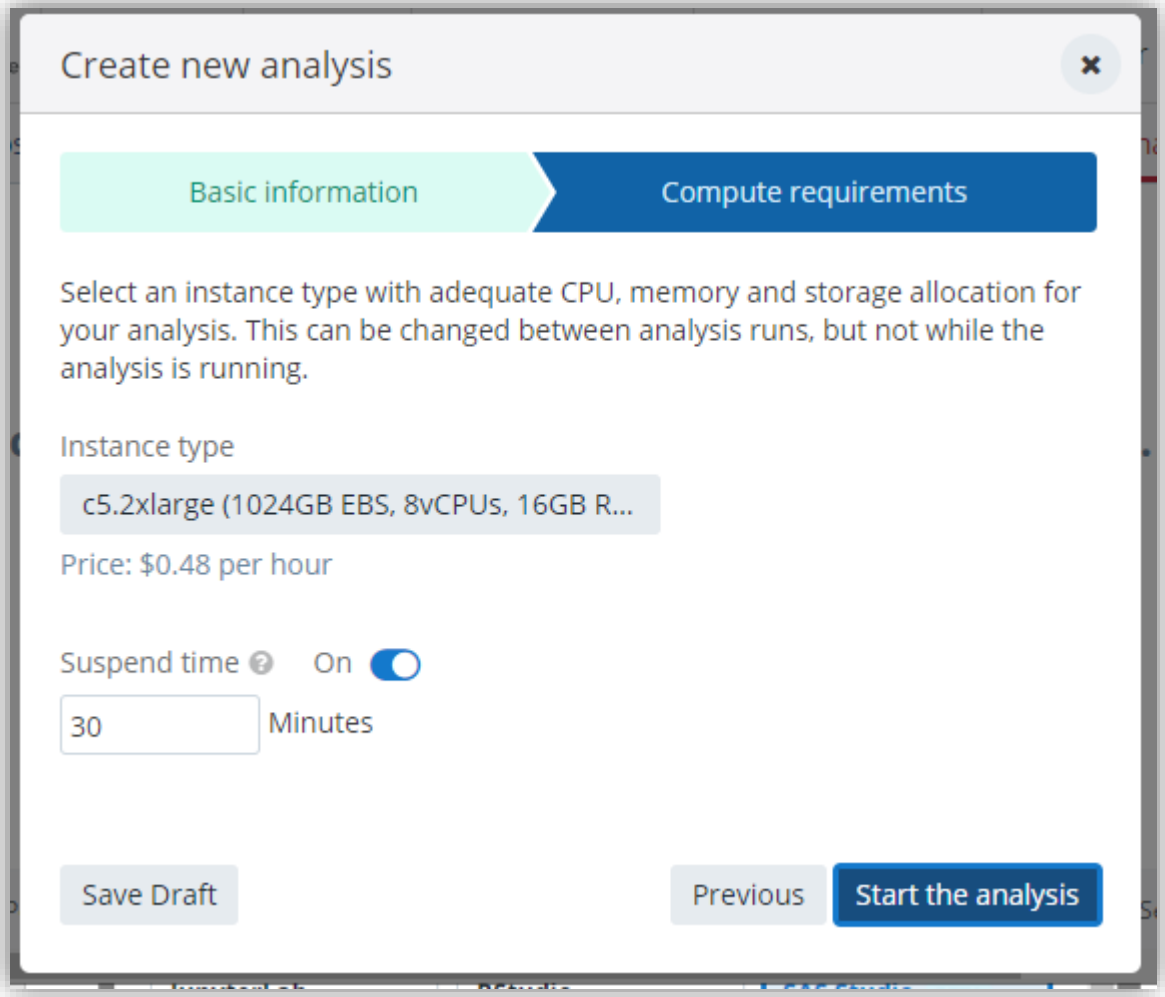


How ar

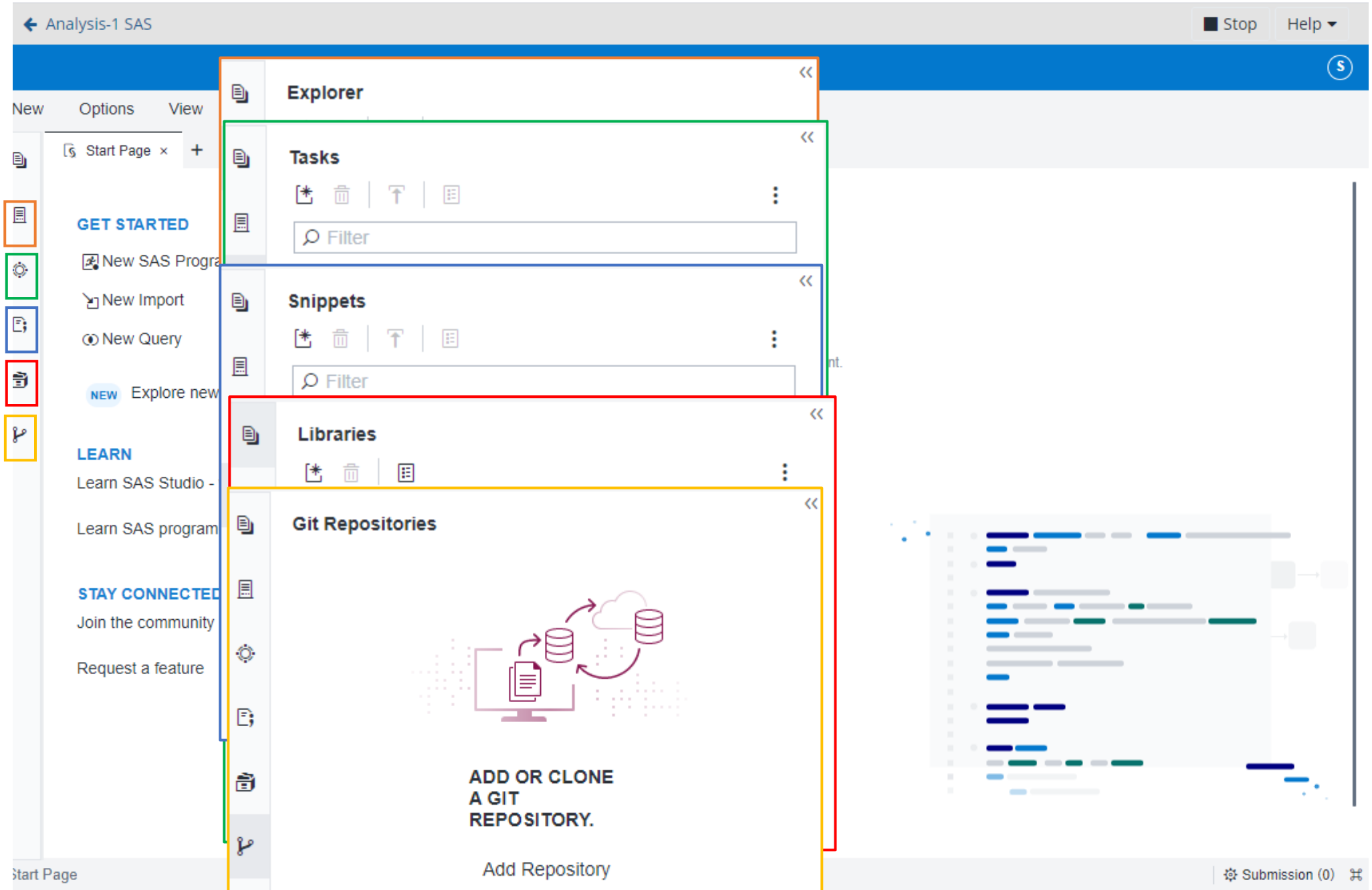

BioData CATALYST
Powered by Seven Bridges
Projects
Data
Public Gallery
Public projects
Developer
pbalte
Dashboard
Files
Apps
Tasks
CONTROLLED
Working group 1
Interactive Analysis
Settings
Notes
Email notifications
Interactive data science and scientific computing across multiple programming languages.
Your analyses will appear here. Learn more
Create your first analysis
eams.
collaborators
ANALYSES
Search
Tasks
Data Cruncher
Your executions will appear here.
Before you start, learn more about them.

Starting analysis in SAS Studio

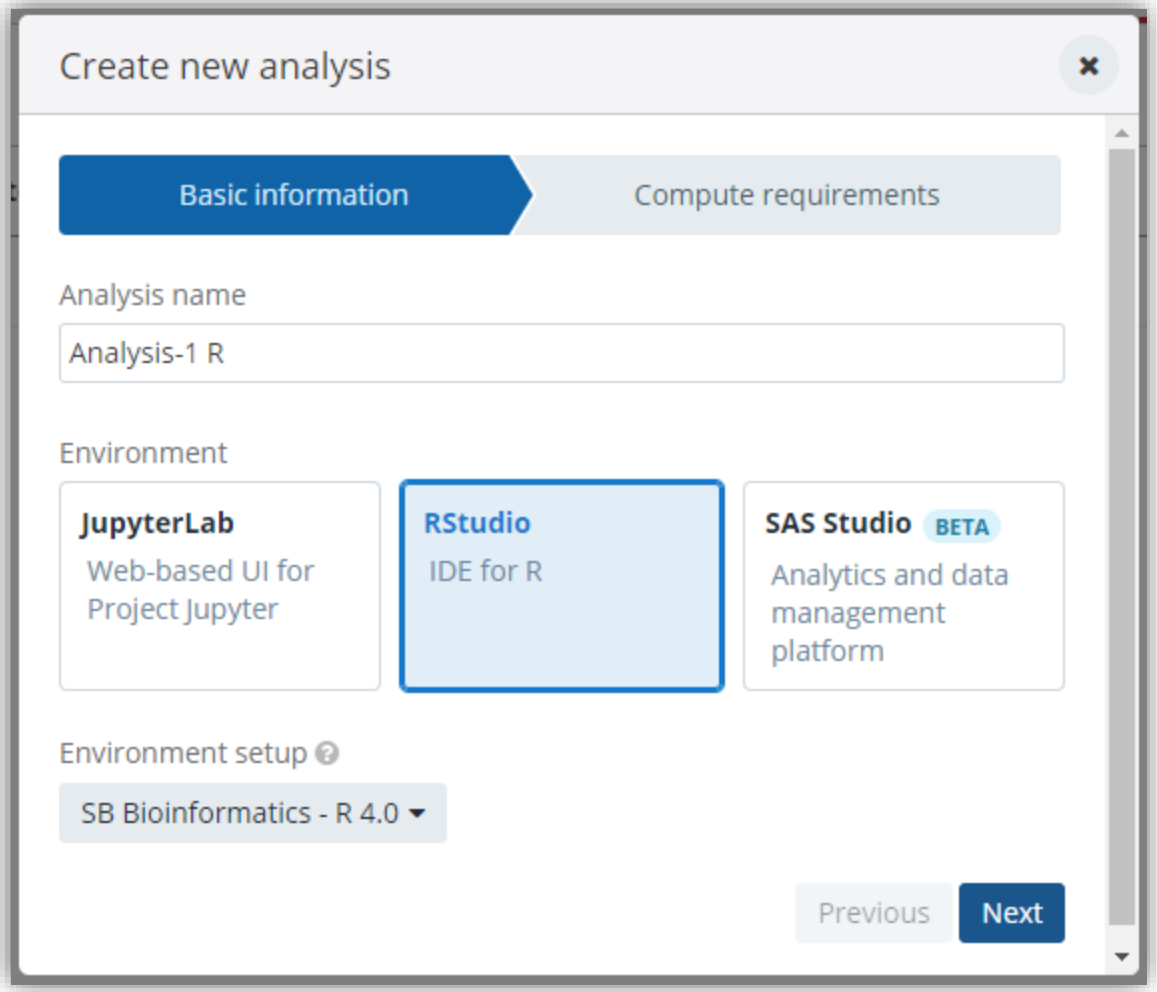


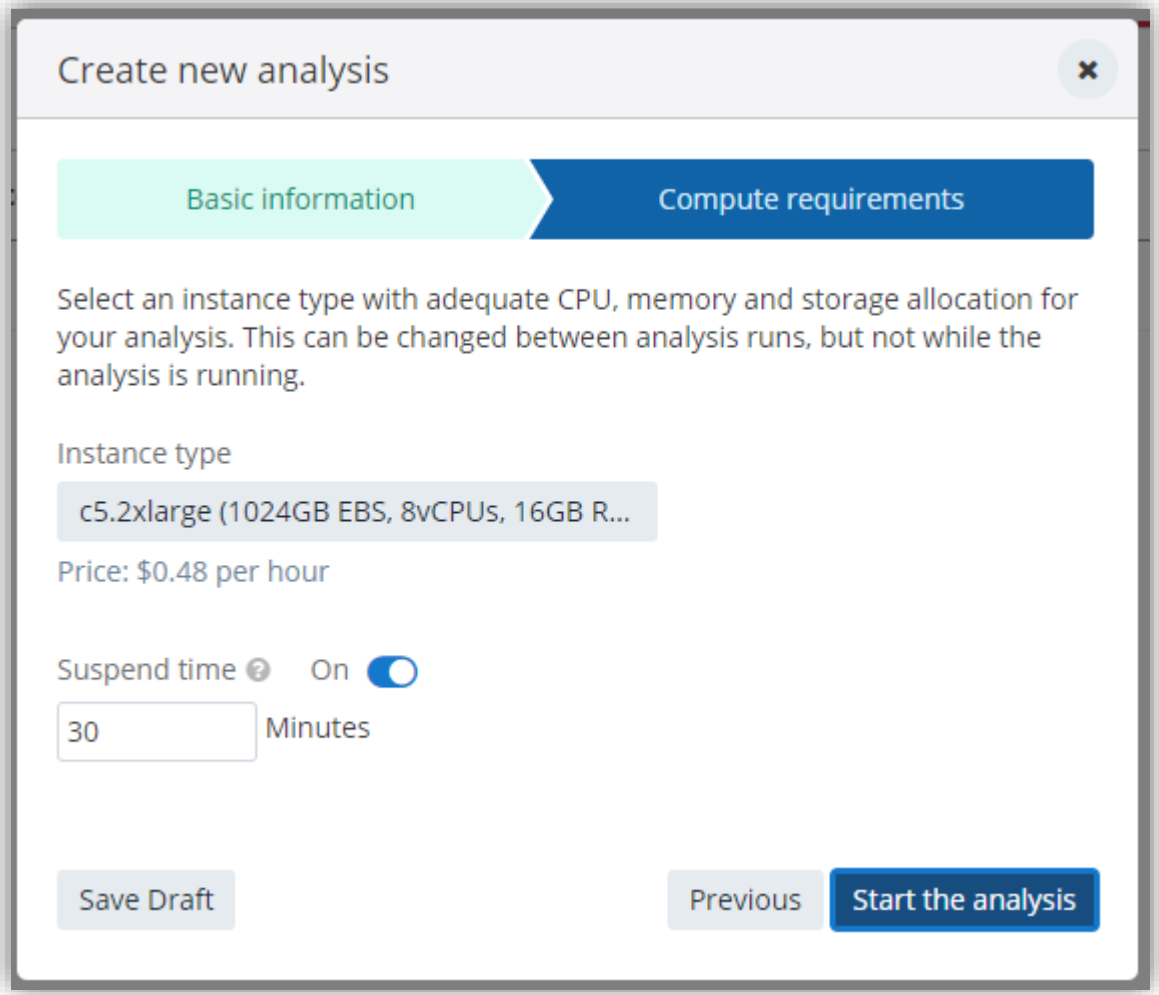


SAS Studio Environment

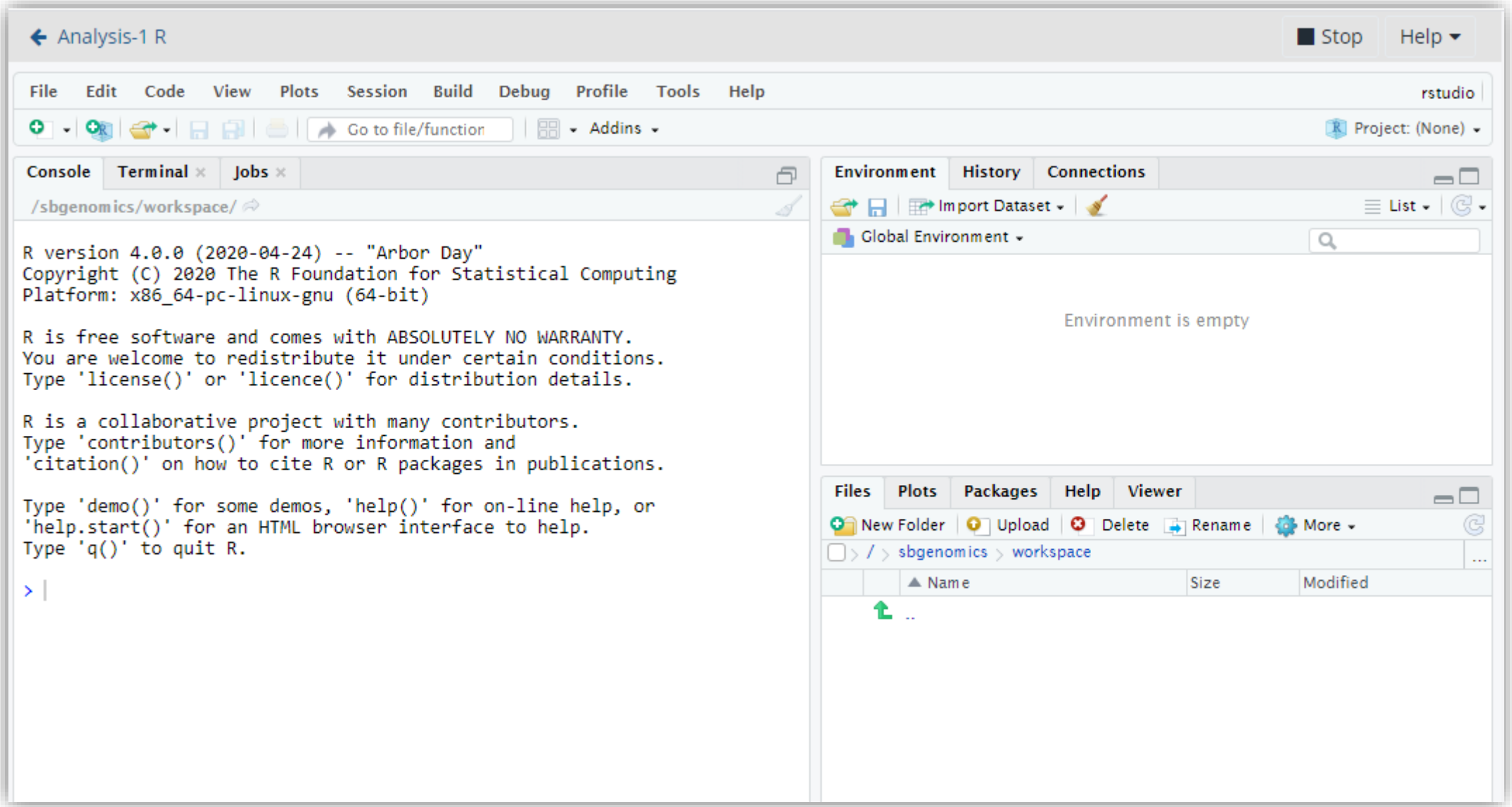


Starting analysis in R Studio





R Studio Environment



Locations of data library in analysis environment

- Data import location: /sbgenomics/project-files/name of the folder
- Data export location: /sbgenomics/output-files
- All exported files are saved in the “Files” of the project
- If desired, exported files can be manually moved to other folders after stopping the analysis

```
libname lib1"/sbgenomics/project-files/DCHC_Datasets";
libname out"/sbgenomics/output-files";
```