



March 15th, 2023 Think-a-Thon

ScHARe

The word "ScHARe" is written in a large, white, bold, sans-serif font. The letter "H" is stylized with a purple arrow pointing to the left and a purple arrow pointing to the right. Behind the letters "H" and "A" are two stylized clouds in shades of orange and yellow.

ScHARe I
Terra Accounts and Workspaces

Sign up for free temporary billing

If you have not filled out the 1-question form on the Think-a-Thon registration confirmation email already, please **provide a Google email address in the chat**

You will be:

- **registered for SchARe**
- **added to a free temporary billing project** that will allow you to run all the Think-a-Thon materials with your instructors

You will be active on this billing project for about 1 day after the Think-a-Thon

If you want to access work-in-progress from the Think-a-Thon after this time, you will need to set up your own billing and copy any of your workspaces to your own billing



Science
collaborative for
Health disparities and
Artificial intelligence bias
Reduction

Sci!ARe



National Institute
on Minority Health
and Health Disparities



Office of
Data Science Strategy



National Institute
of Nursing Research

Thank you

NIMHD
Dr. Eliseo
Perez-Stable

ODSS
Dr. Susan
Gregurick

NIH/OD
Dr. Larry
Tabak

NIMHD OCPL
RLA

BioTeam
STRIDES
Terra
Broad Institute

SIDEM

CCDE Working
Group

Outline

- 5' Introduction and setup
 - Experience poll
- 15' ScHARe and Terra overview
 - Interest poll
- 20' Registration and Account
- 30' Billing and Costs
- 20' Workspaces and Permissions
- 30' Notebooks and Environment

Experience poll

Please check your level of experience with the following:

	None	Some	Proficient	Expert
Python	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
R	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cloud computing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Terra	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Health disparities research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Health outcomes research	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Algorithmic bias mitigation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The logo for ScHARe features the text 'ScHARe' in a white, bold, sans-serif font. The letters 'H' and 'A' are partially obscured by a stylized orange and yellow cloud. A purple double-headed arrow is positioned horizontally across the middle of the 'H' and 'A' characters. The entire logo is set against a dark blue background and has a faint, semi-transparent reflection below it.

ScHARe

Part I

ScHARe and Terra overview



SchARE is a **cloud-based population science data platform** designed to accelerate research in health disparities, health care outcomes, and artificial intelligence (AI) bias mitigation strategies

The platform offers researchers at all career levels and disciplines:

- Access to centralized social determinants of health and other social science **datasets**
- The ability to **collaborate** as they apply AI, machine learning, and other advanced analytical techniques to these datasets in a **secure setting**





SchARE aims to fill **three critical gaps**:

- Foster research collaborations and increase participation of **women and underrepresented populations with health disparities** in data science
- Leverage social science (SDoH) **research opportunities** afforded by Big Data and cloud computing
- Advance **AI bias mitigation** strategies and **ethical inquiry** by increasing the use of diverse eyes and skills



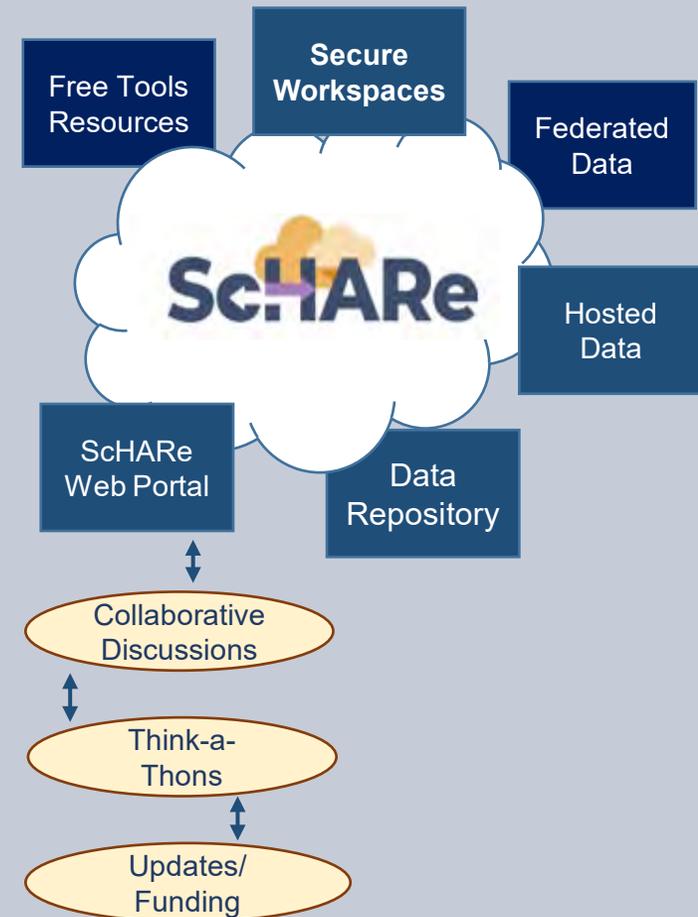
SchARE Components

SchARE co-localizes within the cloud:

- **Datasets** (including social determinants of health and social science data) relevant to minority health, health disparities, and health care outcomes research
- **Data repository** to comply with the required hosting, managing, and sharing of data from NIMHD- and NINR-funded research programs
- **Computational capabilities and secure, collaborative workspaces** for students and all career level researchers
- **Tools for collaboratively evaluating and mitigating biases** associated with datasets and algorithms utilized to inform healthcare and policy decisions

Frameworks: Google Platform, Terra Interface, GitHub, NIMHD Web SchARE Portal

Intramural & Extramural Resource



nimhd.nih.gov/schare

ScHARe Data Ecosystem

Researchers can access, link, analyze, and export a **wealth of datasets** within and across platforms relevant to research about health disparities, health care outcomes and bias mitigation, including:

- **Google Cloud Public Datasets:** publicly accessible, federated, de-identified datasets hosted by Google through the Google Cloud Public Dataset Program
Example: *American Community Survey (ACS)*
- **ScHARe Hosted Public Datasets:** publicly accessible, de-identified datasets hosted by ScHARe
Example: *Behavioral Risk Factor Surveillance System (BRFSS)*
- **Funded Datasets on ScHARe:** publicly accessible and controlled-access, funded program/project datasets using Core Common Data Elements shared by NIH grantees and intramural investigators to comply with the NIH Data Sharing Policy
Examples: *Jackson Heart Study (JHS); Extramural Grant Data; Intramural Project Data*

Access to social and behavioral datasets

- The cloud offers access to vast repositories of data, and enables mapping and linking across sources
- Extremely large datasets are statistically analyzed to gain detailed insights, often using AI and substantial computer-processing power
- Datasets can be linked together to see how patterns in one domain affect other areas
- Data can be structured into fixed fields or unstructured as free-flowing information
- Increase use of Dark Data



ScHARe Data Ecosystem will offer access to **300+ datasets**, including:

- Google Cloud Public Datasets
- ScHARe Hosted Public Datasets
 - American Community Survey
 - U.S. Census
 - Social Vulnerability Index
 - Food Access Research Atlas
 - Medical Expenditure Panel Survey
 - National Environmental Public Health Tracking Network
 - Behavioral Risk Factor Surveillance System
- Repository for Funded Datasets on ScHARe – Compliance with NIH Data Sharing Policy

Data organization



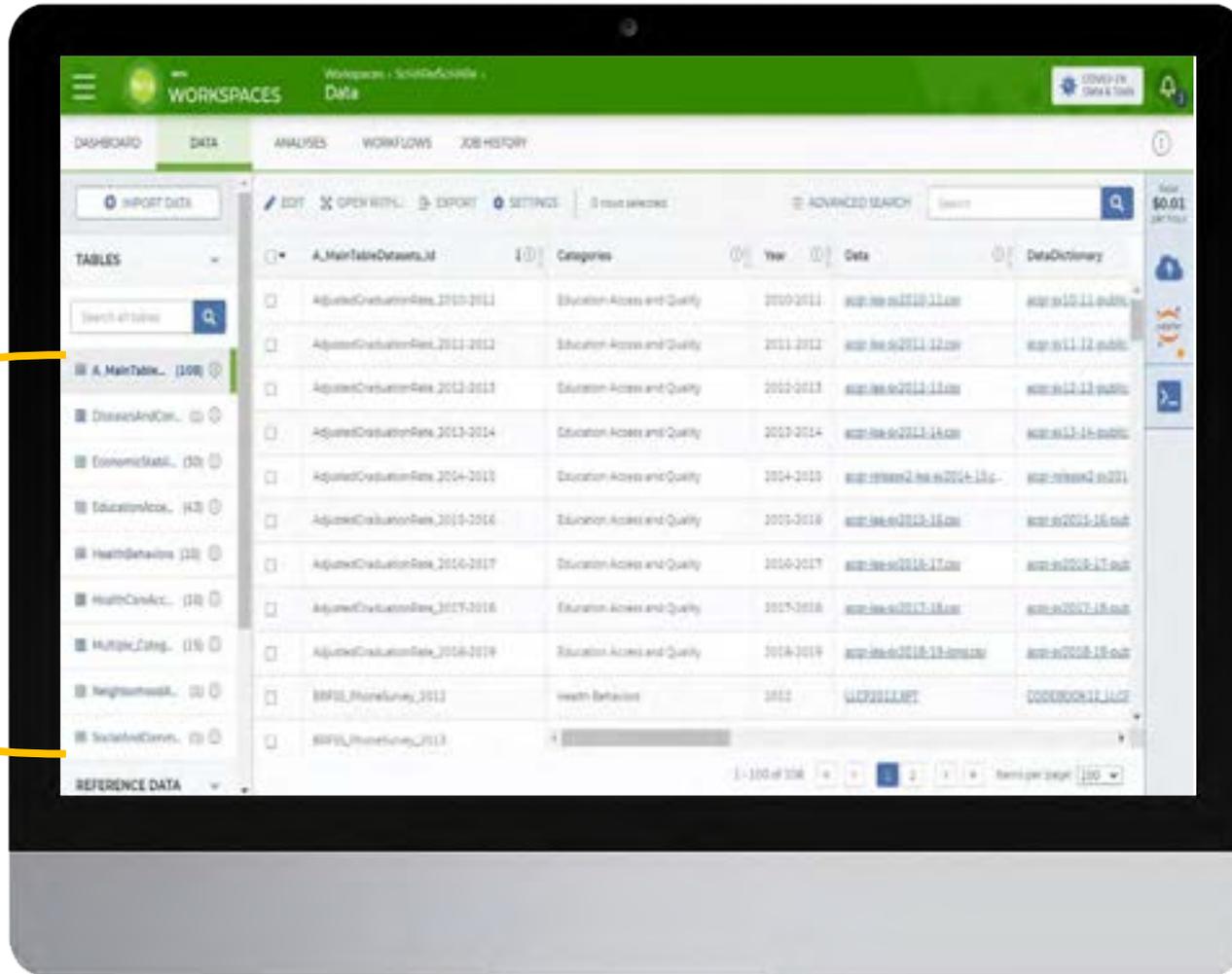
On ScHARe, datasets are categorized by content based on the CDC **Social Determinants of Health** categories:

1. Economic Stability
2. Education Access and Quality
3. Health Care Access and Quality
4. Neighborhood and Built Environment
5. Social and Community Context

with the addition of:

- **Health Behaviors**
- **Diseases and Conditions**

Users will be able to **map and link** across datasets



Cloud computing strategies

- AI/ML can analyze massive amounts of data properly without human intervention
- Terra, standalone or in conjunction with Google Cloud Platform's Vertex AI, can support AI & ML-based analyses
- Workflows (pipelines) are a series of steps performed by a compute engine for bulk analysis
- Computing environments can be customized or standardized (using a custom Docker Image or a startup script)



SciARe

- Uses **workflows** in Workflow Description Language (WDL), a language easy for humans to read, for batch processing data.
- **Python and R**, including libraries most commonly used
- Enables **customization** of computing environments to ensure everyone in your group is using the same software
- **Big query** and **Tensorflow** access for advanced machine learning
- Enables researchers to create interactive **Jupyter notebooks** (documents that contain live code) and share data, analyses and results with their collaborators in real time
- For novice users, integration with **SAS** is planned

Cloud security

- The Terra system has been granted Authority to Operate as a FISMA Moderate impact system and is FedRAMP authorized
- NIST framework NIST-800-53 Rev 4 Moderate, to achieve compliance with industry-accepted security standards



- **Secure Workspaces**
- **Duos** used for organization and security
- **RAS** single-sign on in future

AI bias mitigation strategies

- Widespread use of AI raises a number of ethical, moral, and legal issues – likely not to go away
- Algorithms often are “black boxes”
- Biases can result from:
 - social/cultural context not considered
 - design limitations
 - data missingness and quality problems
 - algorithm development and model training
 - implementation
- If not rectified, biases may result in decisions that lead to discrimination, unequitable healthcare, and/or health disparities
- Lack of diverse perspectives - Populations with health disparities are underrepresented in data science
- Guidelines and recommendations emerging from HHS, NIST, White House, etc.



Critical thinking can rectify AI biases.

SchARE was created to:

- foster participation of **populations with health disparities in data science**
- promote the collaborative identification of **bias mitigation strategies across the continuum**
- create a **culture of ethical inquiry** and critical thinking whenever AI is utilized
- build **community confidence** in implementation approaches
- focus on **implementation of AI bias** guidelines and recommendations

SeHARe Data Repository

CORE COMMON DATA ELEMENTS

**NOVEL CDE FOCUSED
REPOSITORY TO FOSTER
INTEROPERABILITY**

**COMPLY WITH DATA SHARING
POLICY - HOST PROJECT DATA**

DATA ECOSYSTEM

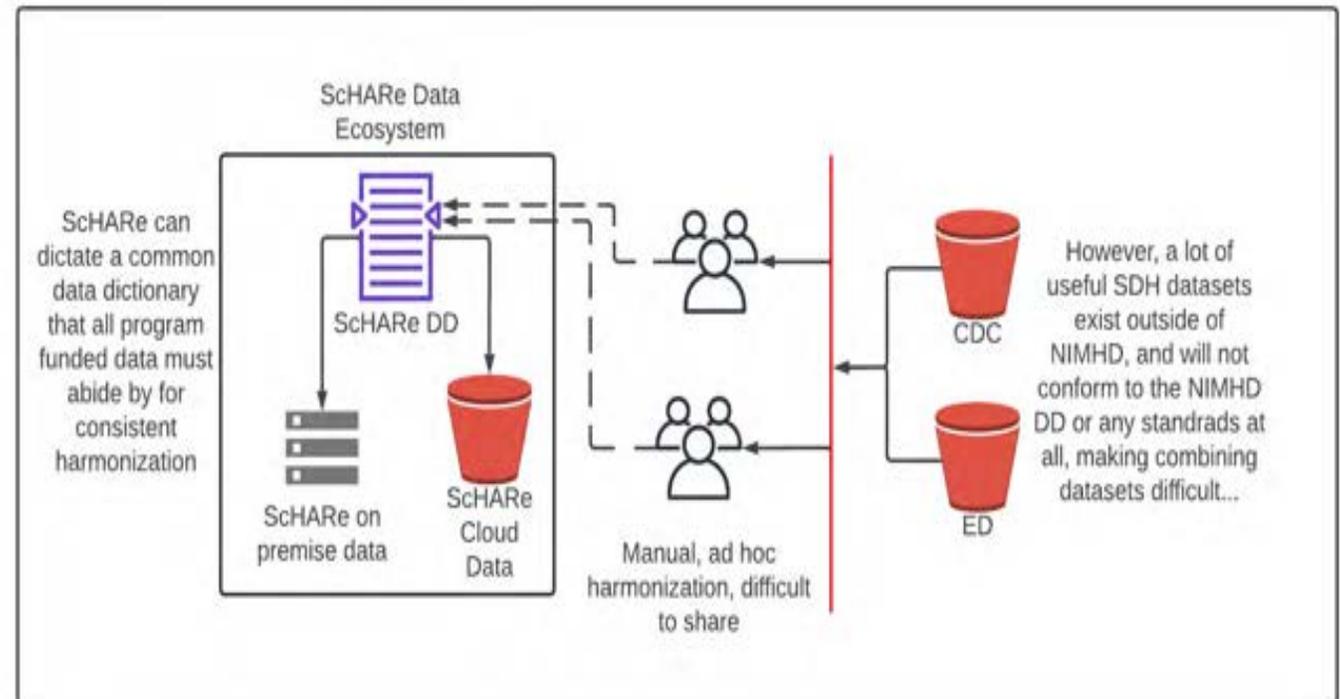
- Map across datasets
- Map across platforms



UPCOMING

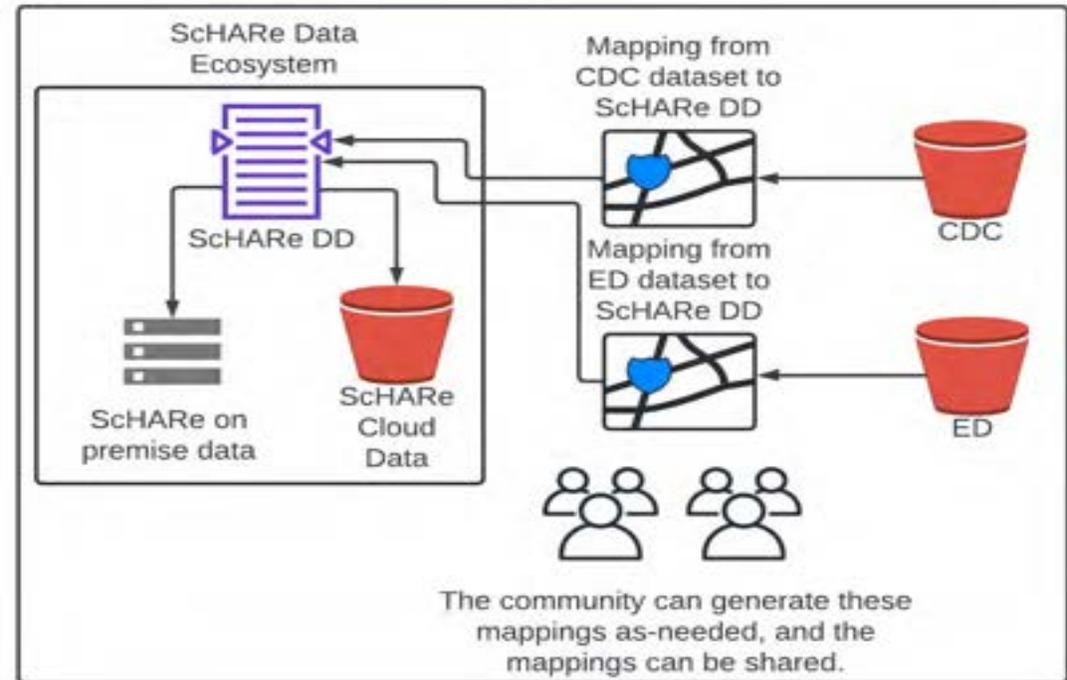
ScHARe Common Data Elements and Harmonization

- Through significant efforts, **data can be harmonized** to a particular data dictionary
- To incorporate external data, users must harmonize in an ad-hoc manner that is **specific to a particular use case**, and requires significant manual effort
- It is **difficult to share these mappings** with others or re-use the ad-hoc harmonization in other projects

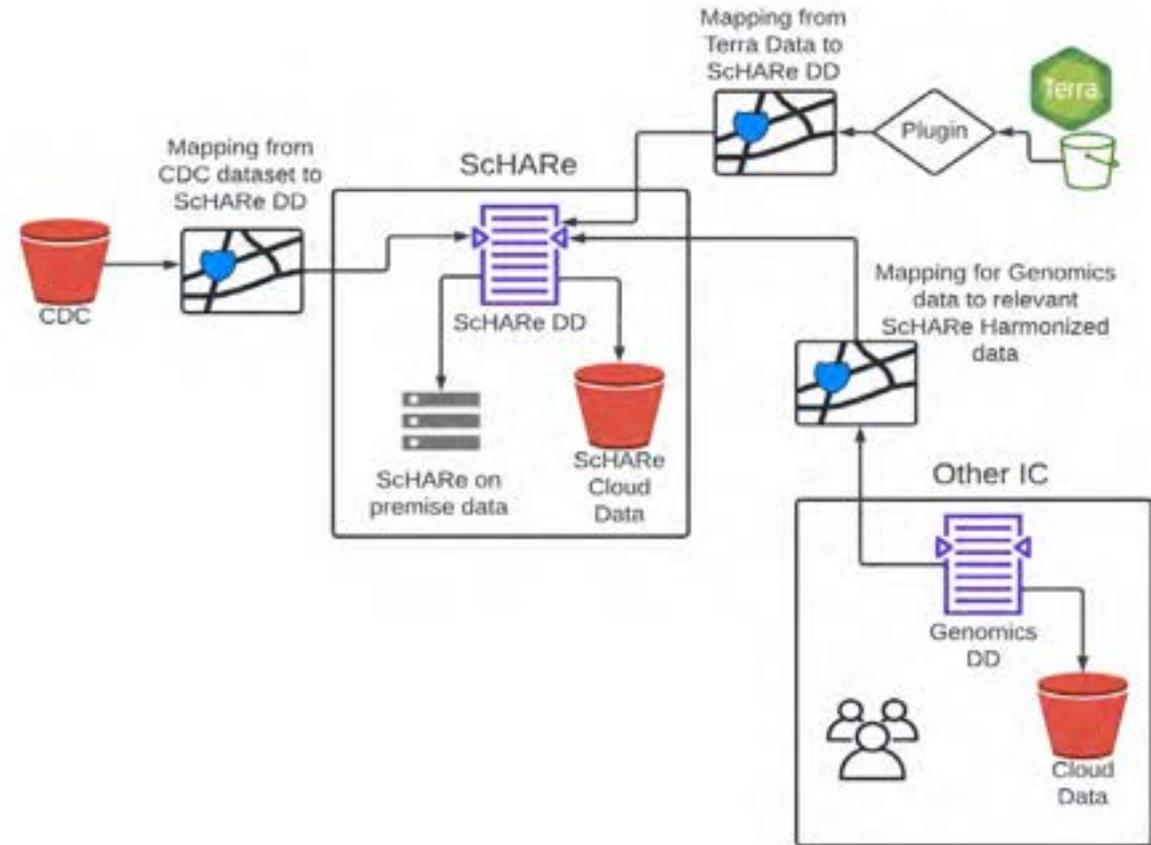


UPCOMING

- **Harmonization should be shared:** data standards cannot be enforced on everyone
- We are building a **mapping system** to relate datasets to data dictionaries and transform them on the fly (we call the result of a transform a **dataview**)
- Mappings can easily be **shared amongst the community:** a user who mapped CDC data to ScHARe's DD can then share it with the community



- By providing **mappings and simple connector plugins** as-needed, we can accommodate new data sources and map between data dictionaries
- This provides the **flexibility to expand the ecosystem** as new standards are developed and adopted
- Current **common data elements** can be incorporated into mappings immediately



The ScHARe logo features the text "ScHARe" in a bold, dark blue font. The letter "H" is stylized with a yellow and orange cloud-like shape above it and a purple arrow pointing to the right below it.

ScHARe

The logo for the All of Us Research Program, with "All of Us" in a large, bold, dark blue font and "RESEARCH PROGRAM" in a smaller, dark blue font below it.

All of Us
RESEARCH PROGRAM

The Terra logo consists of a green hexagonal shape with a white border, containing the word "Terra" in a white, sans-serif font.

Terra

The AnVIL logo features a blue DNA double helix icon to the left of the text "AnVIL" in a bold, blue, sans-serif font.

AnVIL

The BioData Catalyst logo includes the text "BioData" in a dark grey font above the word "CATALYST" in a white, bold, sans-serif font, which is set against a red arrow-shaped background pointing to the right.

BioData
CATALYST

This creates an extraordinary opportunity for **high-impact collaborations** across platforms

Learning how to use Terra on ScHARe will open up **a world of possibilities**, giving you access to an interdisciplinary wealth of datasets and resources

Interest poll

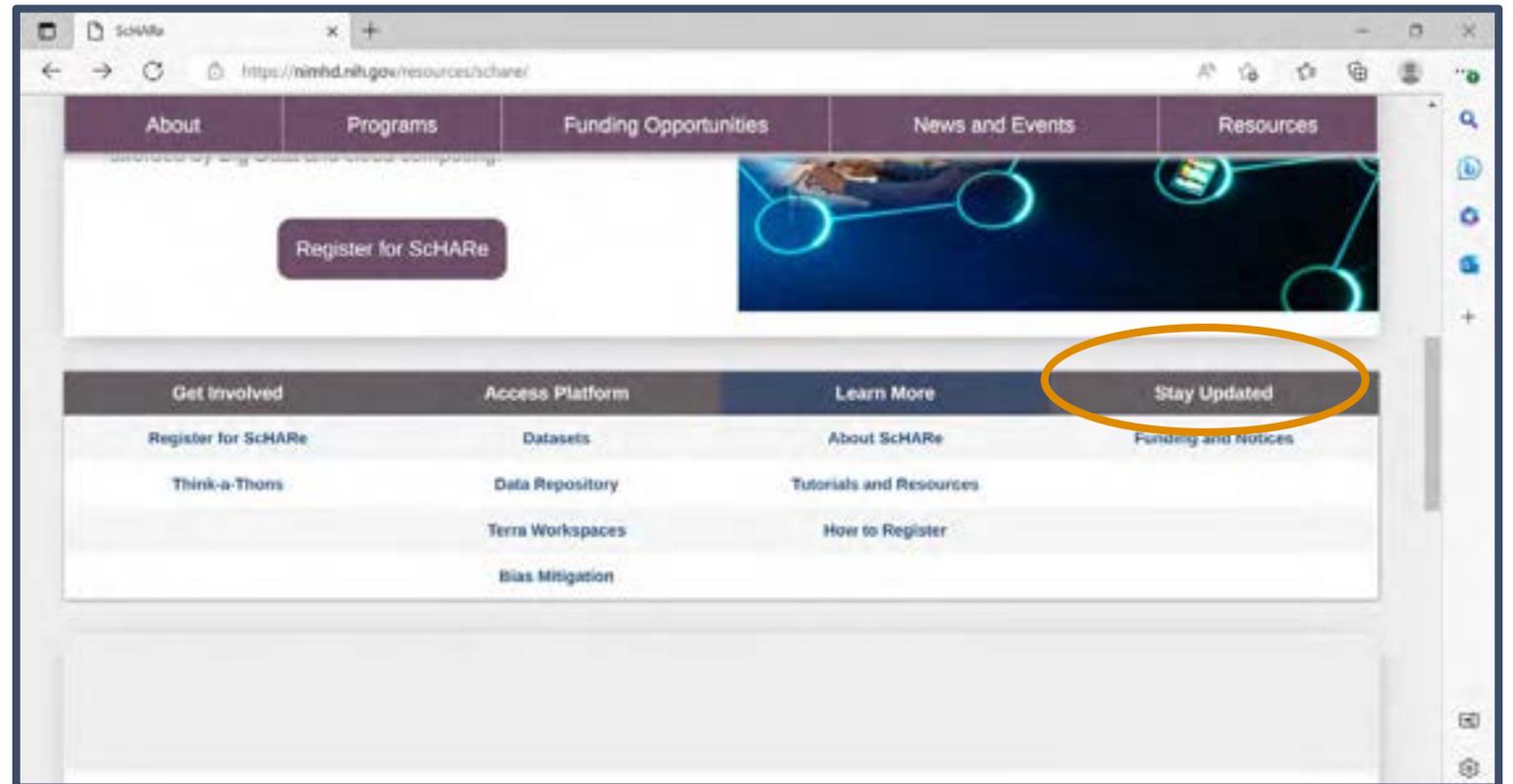
I am interested in (check all that apply):

- Learning about Health Disparities and Health Outcomes research to apply my data science skills
- Conducting my own research using AI/cloud computing and publishing papers
- Connecting with new collaborators to conduct research using AI/cloud computing and publish papers
- Learning to use AI tools and cloud computing to gain new skills for research using Big Data
- Learning cloud computing resources to implement my own cloud
- Developing bias mitigation and ethical AI strategies
- Other

Two Ways to Sign up for ScHARe News



Scannable from your
screen!

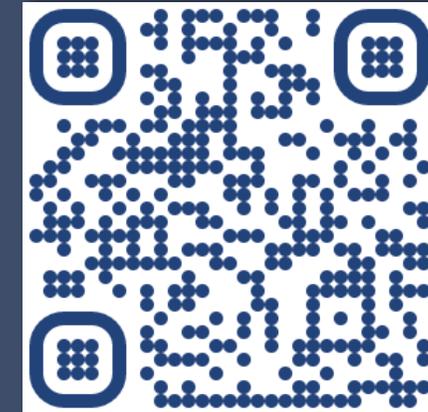


nimhd.nih.gov/schare

SchARE Think-a-Thons

- Monthly sessions (2 hours)
- Instructional/interactive
- Designed for new and experienced users
- Research & analytic teams to:
 - Conduct health disparities, health outcomes, bias mitigation research
 - Analyze/create tools for bias mitigation
- Publications from team collaboration
- Networking
- Mentoring and coaching

Register:



bit.ly/think-a-thons

Instructional

Order	Date/Time	Substitution Number, Title, Description, and Goals
1	February 15, 2023 2:30-4:30 pm	<p>Title: Artificial Intelligence and Cloud Computing 101</p> <p>Description: An introduction to the artificial intelligence and cloud computing concepts behind the SchARE platform: fundamental terminology, cloud architecture, storage and providers, data structures and management, technology benefits and concerns in the research collaboration world.</p>
1	March 15, 2023 2:30-4:30 pm	<p>Title: SchARE 1 - Research in the Cloud Implementation Strategies</p> <p>Description: An overview of active research collaboration platforms, success stories, best initiatives and strategies for low-cost cloud development, grant writing 101 for cloud implementation projects, SchARE mission and vision.</p>
1	April 18, 2023 2:30-4:30 pm	<p>Title: SchARE 2 - Accounts, Workspaces, and Analysis</p> <p>Description: An inside look at SchARE's Terra instance: general introduction and features; how to create and configure an account and set up billing; how to create a workspace and set the appropriate permissions; how to clone or create and run a notebook; data and workflow analysis.</p>

Research teams

1	June 11, 2023 2:30-4:30 pm	<p>Title: Data Science Projects 1 - Health Disparities and Individual SDOH</p> <p>Description: Exploring the impact of individual Social Determinants of Health on Health outcomes: a hands-on session for researchers and students at all levels interested in collaborating on SchARE to develop innovative research questions and projects leading to publications.</p>
2	July 19, 2023 2:30-4:30 pm	<p>Title: Data Science Projects 2 - Health Disparities and Structural SDOH</p> <p>Description: Assessing the impact of structural Social Determinants of Health on Health outcomes: a hands-on session for researchers and students at all levels interested in collaborating on SchARE to develop innovative research questions and projects leading to publications.</p>
2	August 16, 2023 2:30-4:30 pm	<p>Title: Data Science Projects 3 - Health Outcomes</p> <p>Description: Investigating the influence of non-clinical factors on disparities in health care delivery: a hands-on session for researchers and students at all levels interested in collaborating on SchARE to develop innovative research questions and projects leading to publications.</p>



SCIARe

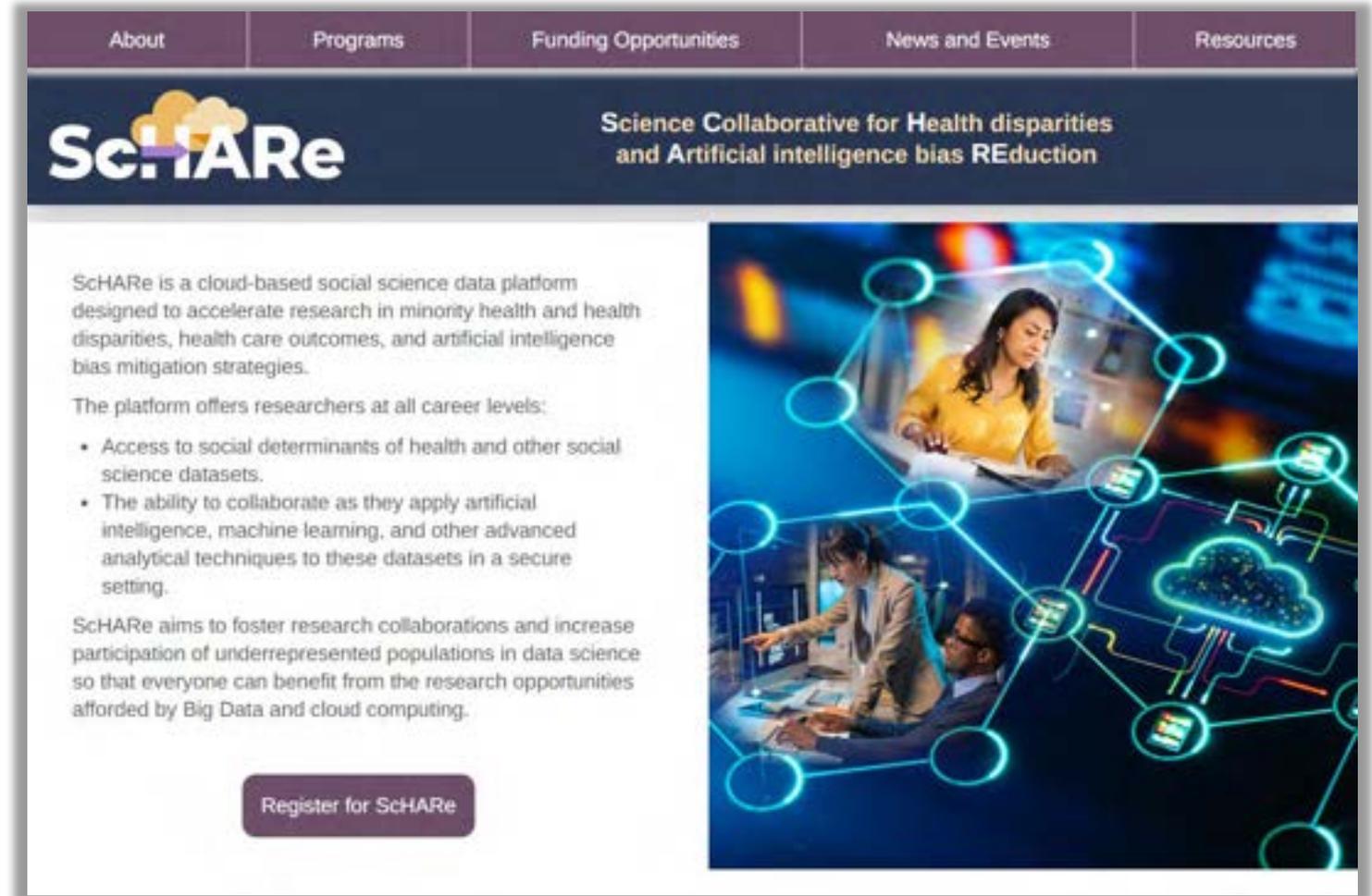
Part II

Registration and Account

Registering for ScHARe

Complete the following steps to register for ScHARe:

1. Visit the ScHARe portal on the NIMHD website:
nimhd.nih.gov/schare
2. Click on the “Register for ScHARe” button



The screenshot shows the ScHARe website homepage. At the top, there is a navigation menu with links for "About", "Programs", "Funding Opportunities", "News and Events", and "Resources". Below the menu is the ScHARe logo and the full name: "Science Collaborative for Health disparities and Artificial intelligence bias REduction". The main content area features a text block describing the platform, a list of benefits, and a "Register for ScHARe" button. To the right of the text is a large image of researchers in a lab setting with a glowing network diagram overlaid.

About Programs Funding Opportunities News and Events Resources

ScHARe

Science Collaborative for Health disparities and Artificial intelligence bias REduction

ScHARe is a cloud-based social science data platform designed to accelerate research in minority health and health disparities, health care outcomes, and artificial intelligence bias mitigation strategies.

The platform offers researchers at all career levels:

- Access to social determinants of health and other social science datasets.
- The ability to collaborate as they apply artificial intelligence, machine learning, and other advanced analytical techniques to these datasets in a secure setting.

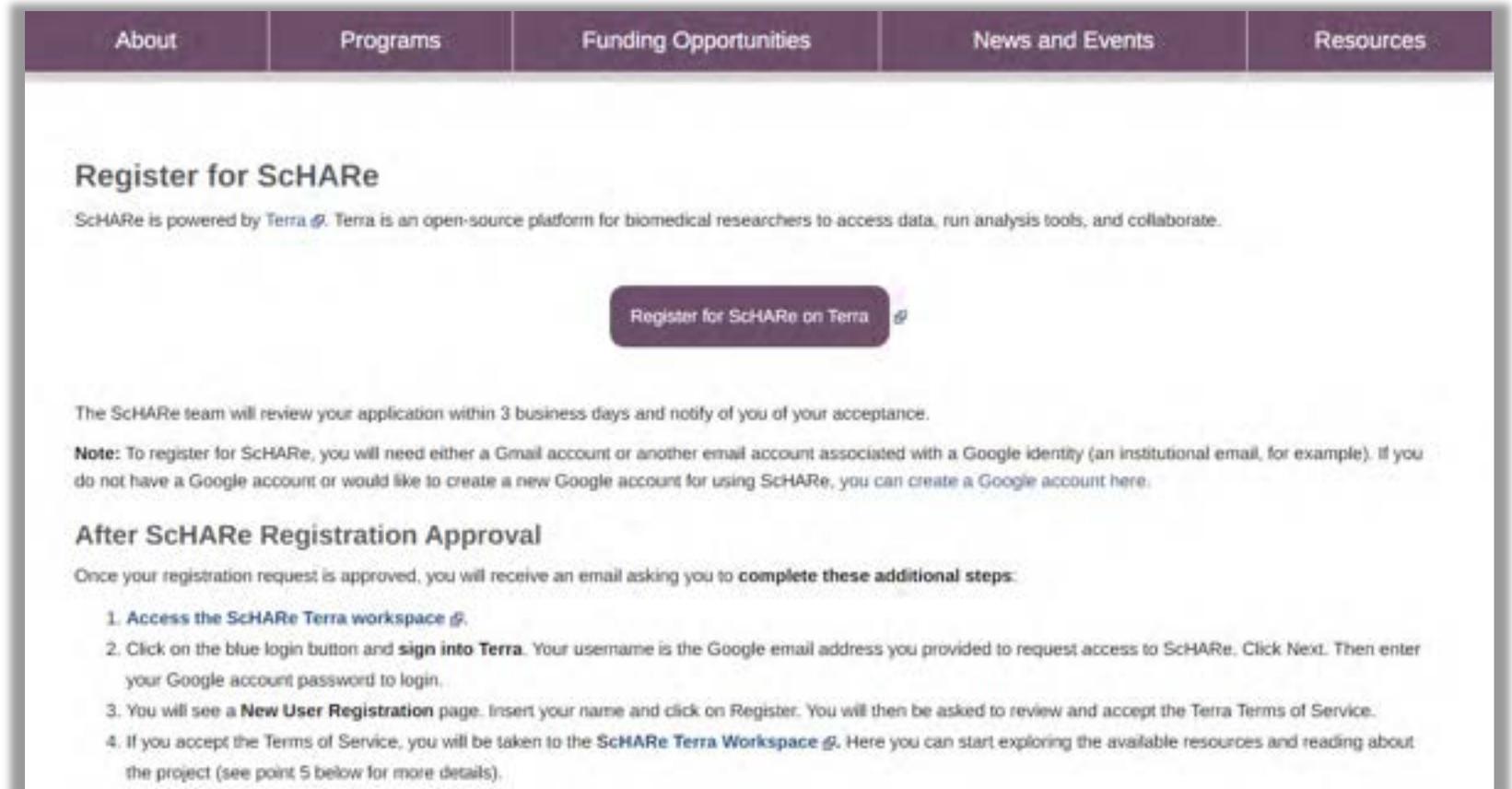
ScHARe aims to foster research collaborations and increase participation of underrepresented populations in data science so that everyone can benefit from the research opportunities afforded by Big Data and cloud computing.

Register for ScHARe

Registering for ScHARe

On the registration page:

Click on the “Register for ScHARe on Terra” button



The screenshot shows the ScHARe registration page on Terra. At the top, there is a navigation bar with five tabs: About, Programs, Funding Opportunities, News and Events, and Resources. Below the navigation bar, the main heading is "Register for ScHARe". Underneath this heading, a sub-heading states: "ScHARe is powered by Terra @. Terra is an open-source platform for biomedical researchers to access data, run analysis tools, and collaborate." A prominent purple button with white text reads "Register for ScHARe on Terra @". Below the button, a paragraph states: "The ScHARe team will review your application within 3 business days and notify of you of your acceptance." A **Note** follows: "To register for ScHARe, you will need either a Gmail account or another email account associated with a Google identity (an institutional email, for example). If you do not have a Google account or would like to create a new Google account for using ScHARe, you can create a Google account here." The next section is titled "After ScHARe Registration Approval" and contains the text: "Once your registration request is approved, you will receive an email asking you to **complete these additional steps**:" followed by a numbered list of four steps.

About Programs Funding Opportunities News and Events Resources

Register for ScHARe

ScHARe is powered by Terra @. Terra is an open-source platform for biomedical researchers to access data, run analysis tools, and collaborate.

Register for ScHARe on Terra @

The ScHARe team will review your application within 3 business days and notify of you of your acceptance.

Note: To register for ScHARe, you will need either a Gmail account or another email account associated with a Google identity (an institutional email, for example). If you do not have a Google account or would like to create a new Google account for using ScHARe, you can create a Google account here.

After ScHARe Registration Approval

Once your registration request is approved, you will receive an email asking you to **complete these additional steps**:

1. Access the ScHARe Terra workspace @.
2. Click on the blue login button and **sign into Terra**. Your username is the Google email address you provided to request access to ScHARe. Click Next. Then enter your Google account password to login.
3. You will see a **New User Registration** page. Insert your name and click on Register. You will then be asked to review and accept the Terra Terms of Service.
4. If you accept the Terms of Service, you will be taken to the **ScHARe Terra Workspace @**. Here you can start exploring the available resources and reading about the project (see point 5 below for more details).

Registering for ScHARe

Complete the registration form

The ScHARe team will:

- review and approve your application
 - send you an email with additional instructions within 3 business days
- Note: you will need a **Gmail account** or another email account (an institutional email, for example) associated with a Google identity. If you do not have it, you can create one here:

bit.ly/3QeUngh



The screenshot shows the top portion of a registration form. At the top, there is a dark blue header with the ScHARe logo (a stylized orange and yellow cloud above the text 'ScHARe') and the title 'Registration Form' in white. Below the header, the form content is on a white background. It starts with a small asterisk and the word 'Required'. The main heading is 'Welcome to ScHARe'. The first paragraph explains that ScHARe is a National Institutes of Health (NIH) project powered by Terra, designed for biomedical researchers to access data and run analyses in secure online spaces. The second paragraph identifies ScHARe as a collaboration between the National Institute on Minority Health and Health Disparities (NIMHD) and the National Institute of Nursing Research (NINR) at the NIH. The third paragraph states that users must fill out the form to gain access to the ScHARe platform on Terra, and that the ScHARe team will review and approve applications within 3 business days. The final line of text indicates that questions marked with an asterisk are required.

Creating a Terra account

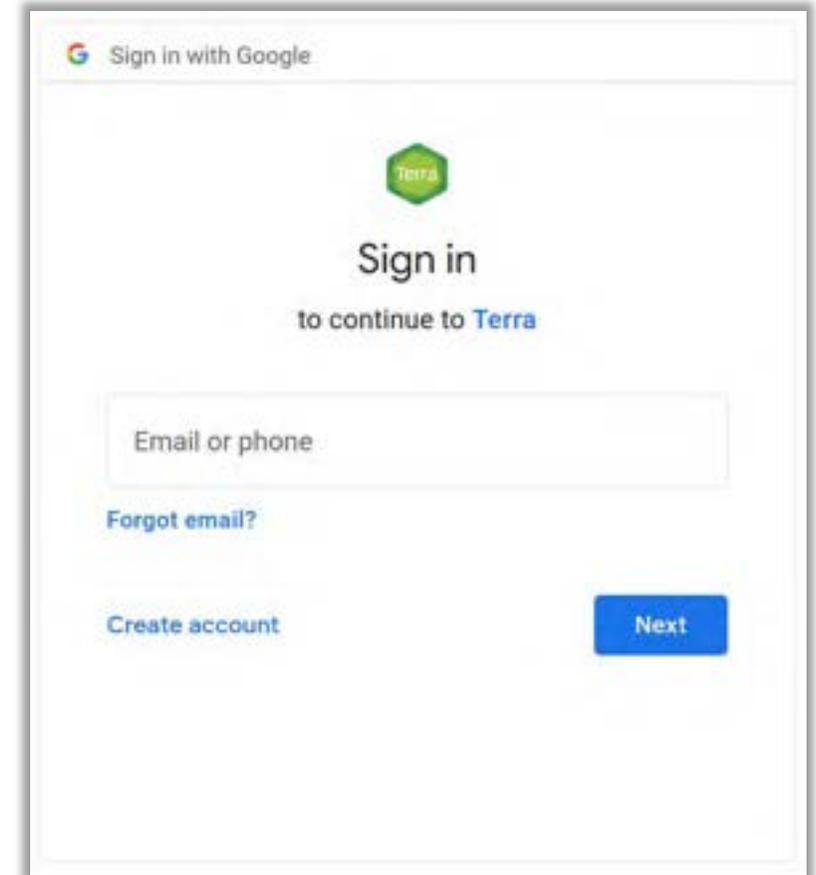
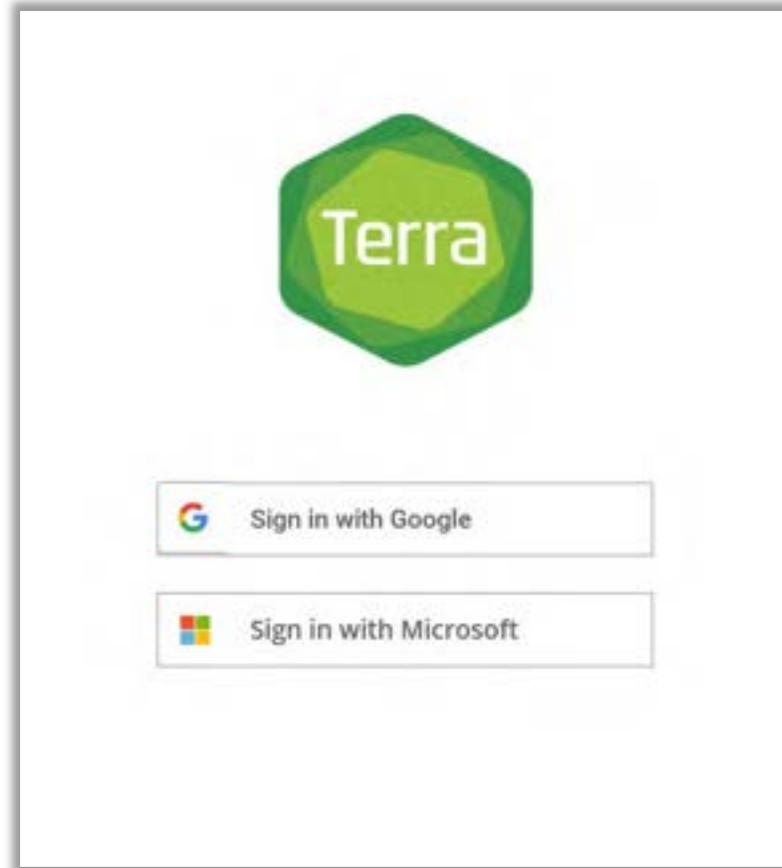
The email you will receive after ScHARe registration approval will ask you to **complete the following steps**:

1. Access the ScHARe Terra workspace at:
bit.ly/access-schare
2. Click on the blue “Log in” button



Creating a Terra account

4. Select “Sign in with Google”
5. Sign into Terra. Your username is the Google email address you provided to request access to ScHARe
6. Click “Next” and enter your Google account password to login



Creating a Terra account

7. You will see a New User Registration page. Insert your name and contact email, then click on “Register”



The screenshot shows the Terra New User Registration page. At the top left is the Terra logo, a green hexagon with the word "Terra" inside. To its right is the word "TERRA" in a large, bold, black sans-serif font. Below the logo and name is the heading "New User Registration". Underneath the heading are three input fields: "First Name *" and "Last Name *" are side-by-side, and "Contact Email for Notifications *" is centered below them. At the bottom of the form are two buttons: "REGISTER" and "CANCEL".

Terra **TERRA**

New User Registration

First Name * Last Name *

Contact Email for Notifications *

REGISTER CANCEL

Creating a Terra account

8. Review and accept the Terra Terms of Service



Terra Terms of Service

Please accept the Terms of Service to continue.

Terra Platform Terms of Service

Last Modified: January 10th, 2023

Effective Date: January 10th, 2023

Certain terms (including capitalized terms) are defined in Section 13 (Definitions).

1. Introduction

Thanks for using Terra! Terra is a platform developed by Verily, Microsoft and The Broad Institute that enables Users to curate and publish datasets, discover and access data,

DECLINE AND SIGN OUT ACCEPT

Access to social and behavioral datasets

- The cloud offers access to vast repositories of data, and enables mapping and linking across sources
- Extremely large datasets are statistically analyzed to gain detailed insights, often using AI and substantial computer-processing power
- Datasets can be linked together to see how patterns in one domain affect other areas
- Data can be structured into fixed fields or unstructured as free-flowing information
- Increase use of Dark Data



ScHARe Data Ecosystem will offer access to **300+ datasets**, including:

- Google Cloud Public Datasets
- ScHARe Hosted Public Datasets
 - American Community Survey
 - U.S. Census
 - Social Vulnerability Index
 - Food Access Research Atlas
 - Medical Expenditure Panel Survey
 - National Environmental Public Health Tracking Network
 - Behavioral Risk Factor Surveillance System
- Repository for Funded Datasets on ScHARe – Compliance with NIH Data Sharing Policy



SCIARe

Part III

Workspaces and Permissions

What is a workspace

Workspaces are the building blocks of Terra - a dedicated space where you and your collaborators can access and organize the same data and tools and run analyses together

They are like **computational sandboxes** with everything you need to complete your project: data, analysis tools, documentation

You can use workspaces to:

- **Link to data in the cloud** for analysis, instead of downloading and storing it yourself
- **Combine data** from different sources in a single table for analysis
- **Keep data organized with integrated spreadsheet-like tables** - no matter where in the cloud the data are stored
- **Visualize and analyze data** in real time using Python and R (and soon, SAS)
- **Find and run bulk analysis tools** (workflows) even if you're not a programming expert
- **Share reproducible analysis results**
- **Collaborate** while maintaining control of your resources

Dashboard

The landing page (i.e., **Dashboard**) is your **project overview** - the questions you're trying to answer, the data and analysis tools you'll use, etc.

The dashboard uses the **Markdown language**, which lets you organize text with headers and include links and additional references



The screenshot shows the ScHARe dashboard interface. At the top, there is a navigation bar with five tabs: DASHBOARD (highlighted in green), DATA, ANALYSES, WORKFLOWS, and JOB HISTORY. Below the navigation bar, the main content area features a header "ABOUT THE WORKSPACE" with a pencil icon. The central focus is the ScHARe logo, where the letters "Sc" and "HARe" are in a dark blue font, and the "H" is replaced by a stylized orange cloud. A purple arrow points from the "H" to the "A". Below the logo, the text "Science Collaborative for Health disparities and AI bias REduction" is displayed. The main body of the page contains a paragraph describing ScHARe as a cloud-based research collaboration platform developed by NIMHD and NINR. It also mentions the program's aim to increase participation of underrepresented populations in data science and cloud computing, and highlights the ScHARe Think-a-Thons webinar series.

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

ABOUT THE WORKSPACE ✎

ScHARe

Science Collaborative for Health disparities and AI bias REduction

[ScHARe](#) (Science Collaborative for Health Disparities and Artificial Intelligence bias Reduction) is a cloud-based research collaboration platform developed by the National Institute on Minority Health and Health Disparities ([NIMHD](#)) and the National Institute of Nursing Research ([NINR](#)).

The aim of the ScHARe program is to increase participation of underrepresented populations in data science and cloud computing so that everyone can benefit from the research opportunities afforded by Big Data.

An integral part of the program is the [ScHARe Think-a-Thons](#) webinar series, conceived to help prepare underrepresented researchers, students, and their collaborators to use the ScHARe platform and learn how to leverage Big Data and cloud computing. Participants will share knowledge and skills, and form cross-disciplinary, multi-level collaborations around innovative research projects that can lead to breakthrough publications.

Dashboard

Workspace details are populated automatically in the right column of the Dashboard

Expandable sections include:

- **Workspace information:** creation date, date last updated, your access level
- **Cloud information:** location of workspace storage, estimated storage cost and size, etc.
- **Workspace owners**
- **Workspace tags:** only visible to owners, tags are useful for searching and indexing

The screenshot displays a dashboard with three main expandable sections:

- WORKSPACE INFORMATION** (expanded):
 - Last Updated: 3/6/2023
 - Creation Date: 1/10/2023
 - Workflow Submissions: 0
 - Access Level: Project Owner
- CLOUD INFORMATION** (expanded):
 - Cloud Name: Google Cloud
 - Location: us us-central1 (Iowa)
 - Google Project ID: terra-d3cb8107
 - Bucket Name: fc-secure-d6e25d73-...
 - Estimated Storage Cost: \$0.43 (Updated on 3/11/2023)
 - Bucket Size: 21.4 GiB (Updated on 3/11/2023)
 - Open bucket in browser
 - Open project in Google Cloud Console
- OWNERS** (collapsed, indicated by a downward arrow and an information icon)

Data

In the **Data** tab, like spreadsheets built right into the workspace, **data tables help keep track of all project data**, no matter where files are stored in the cloud

- In the ScHARe workspace, click on the Data tab
- Under Tables, you will see a list of dataset categories
- If you click on a category, you will see a list of relevant datasets
- Scroll to the right to learn more about each dataset

More information will be provided in the next Think-a-Thons

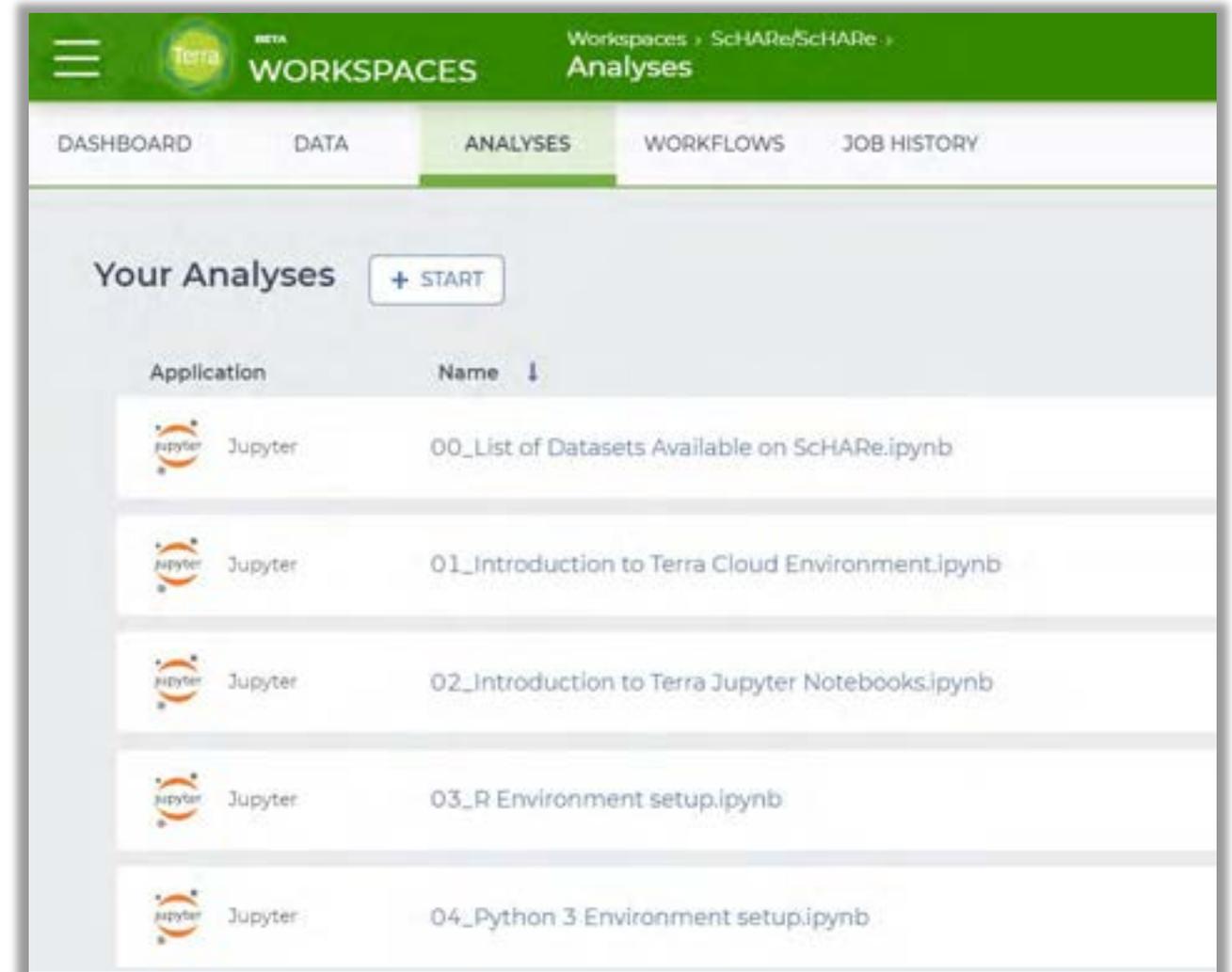
The screenshot shows the ScHARe workspace interface. At the top, there's a green header with the 'Terta' logo and 'WORKSPACES' text. Below the header, there are navigation tabs: DASHBOARD, DATA (selected), ANALYSES, WORKFLOWS, and JOB HISTORY. The main area is divided into two sections. On the left, under 'TABLES', there's a search bar and a list of dataset categories: A_MainTableDatasets (118), DiseaseAndConditions (1), EconomicStability (30), EducationAccessAndQuality (47) (highlighted), EducationAccessAndQuality (47 rows), HealthCareAccessAndQuality (10), MultipleCategories (15), NeighborhoodAndBuiltEnvironment (10), and SocialAndCommunityContext (4). On the right, there's a table with columns 'EducationAccessAndQuality_Id' and 'Categories'. The table contains rows for 'AdjustedGraduationRate' for various years (2010-2011 to 2018-2019) and 'ECPP_EarlyChildhoodProgramParticip...'. The interface also includes an 'IMPORT DATA' button, 'EDIT', 'OPEN WITH...', 'EXPORT', and 'SETTINGS' options, and a '0 rows selected' indicator.

Analyses

In the **Analyses** tab, you can interact with the data and perform/share analyses using Python or R

- In the ScHARe workspace, click on the **Analyses** tab
- This tab contains instructional notebooks and analysis tutorials on how to use the ScHARe resources in the Terra environment

More information on our instructional notebooks will be provided later



The screenshot displays the Terra WORKSPACES interface. The top navigation bar is green and contains the Terra logo, the word "WORKSPACES", and the current workspace path "Workspaces > ScHARe/ScHARe > Analyses". Below this is a secondary navigation bar with tabs for "DASHBOARD", "DATA", "ANALYSES" (which is highlighted), "WORKFLOWS", and "JOB HISTORY". The main content area is titled "Your Analyses" and includes a "+ START" button. Below this is a table listing several Jupyter notebooks:

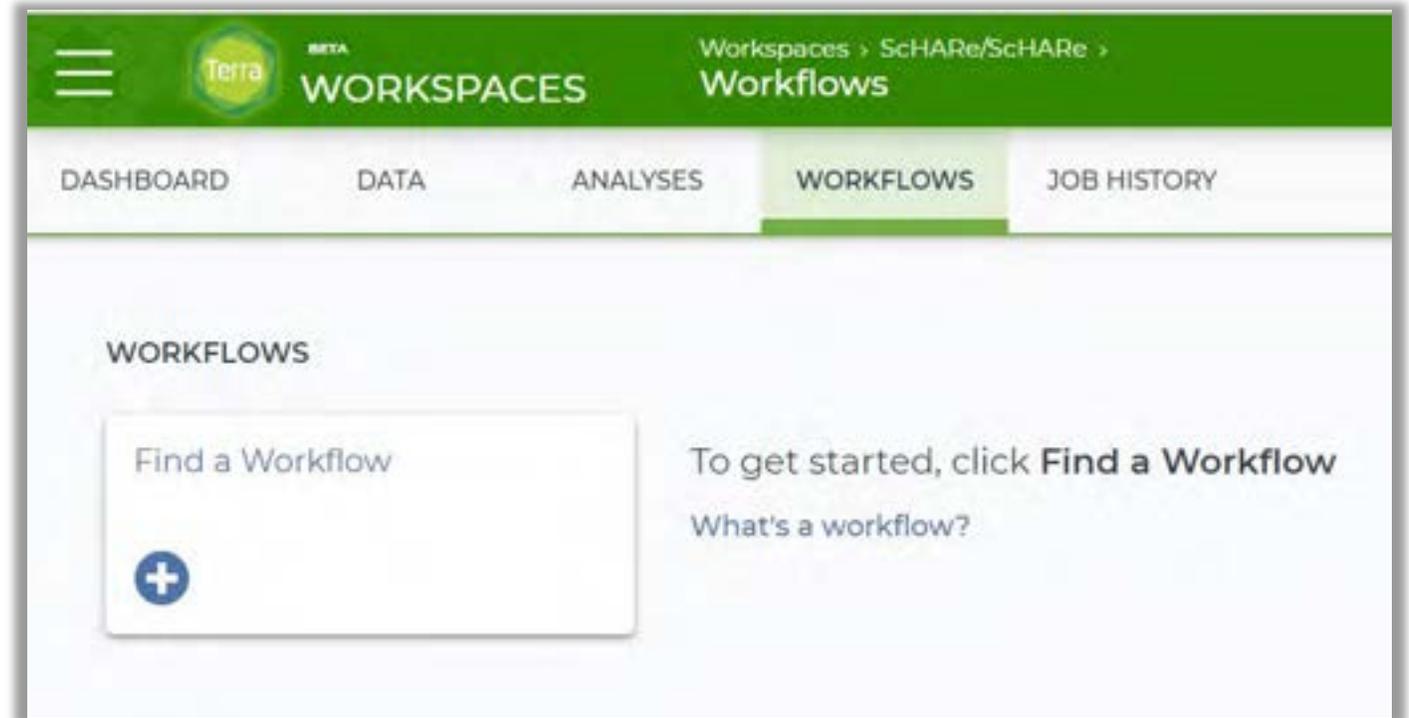
Application	Name ↓
Jupyter	00_List of Datasets Available on ScHARe.ipynb
Jupyter	01_Introduction to Terra Cloud Environment.ipynb
Jupyter	02_Introduction to Terra Jupyter Notebooks.ipynb
Jupyter	03_R Environment setup.ipynb
Jupyter	04_Python 3 Environment setup.ipynb

Workflows

In the **Workflows** tab, you will find workflows for bulk analyses

- These are the sorts of repetitive analyses that can be automated
- Workflows in Terra are written in the human-readable Workflow Description Language (WDL)

More information will be provided in the next Think-a-Thons



Creating your first workspace

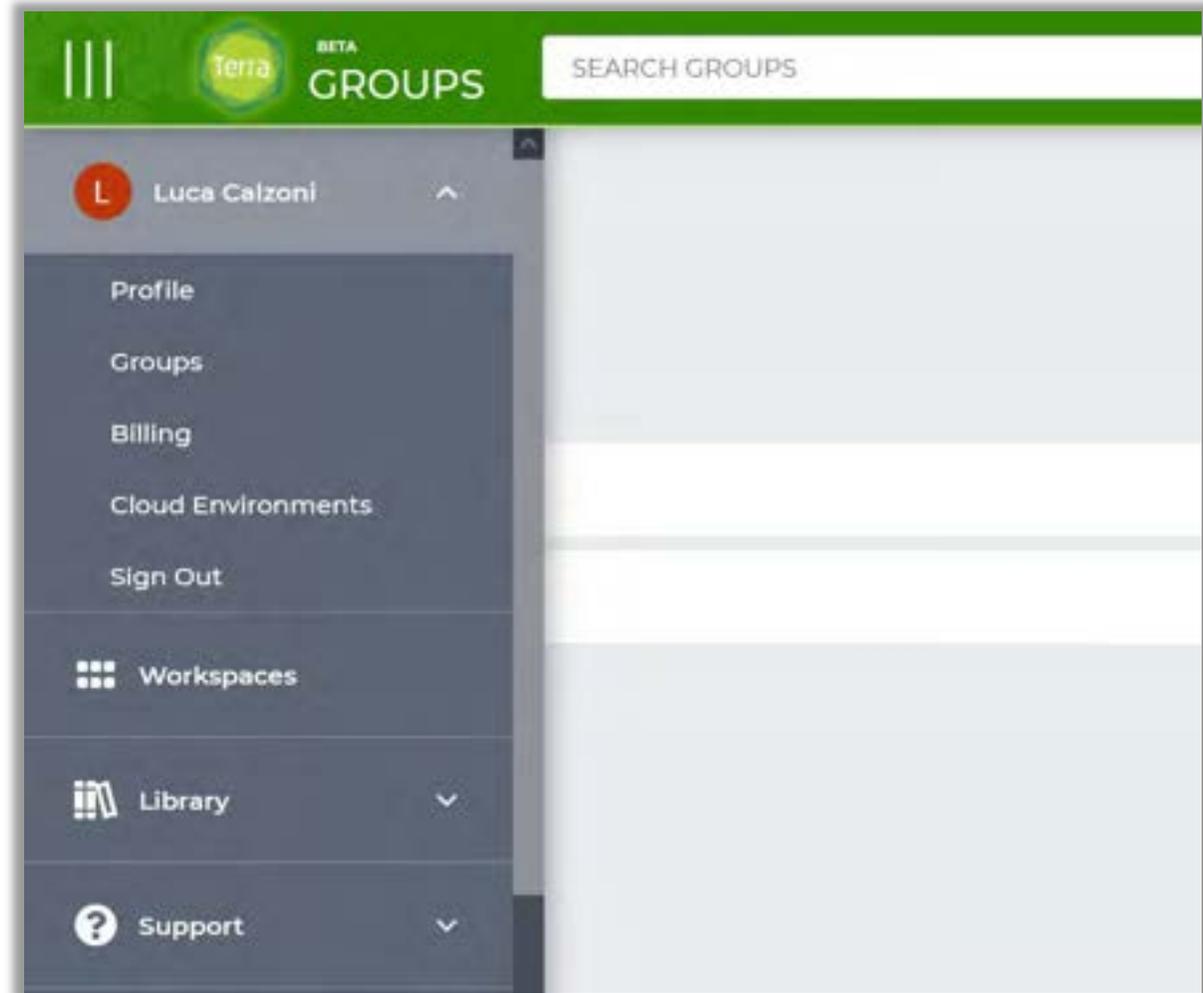
Let's create your first Terra workspace!

For the purpose of this tutorial, let's assume that you intend to create a workspace that will allow you to work with two groups of collaborators:

- **Group 1 – Internal collaborators:** researchers in your lab, who must be able to access your data, perform computations, and work with you to write the collaborative notebooks you plan to use to share your results with the scientific community
- **Group 2 – External collaborators:** researchers at another institution, who you want to be able to see your data, notebooks and analyses, but without the possibility of modifying them

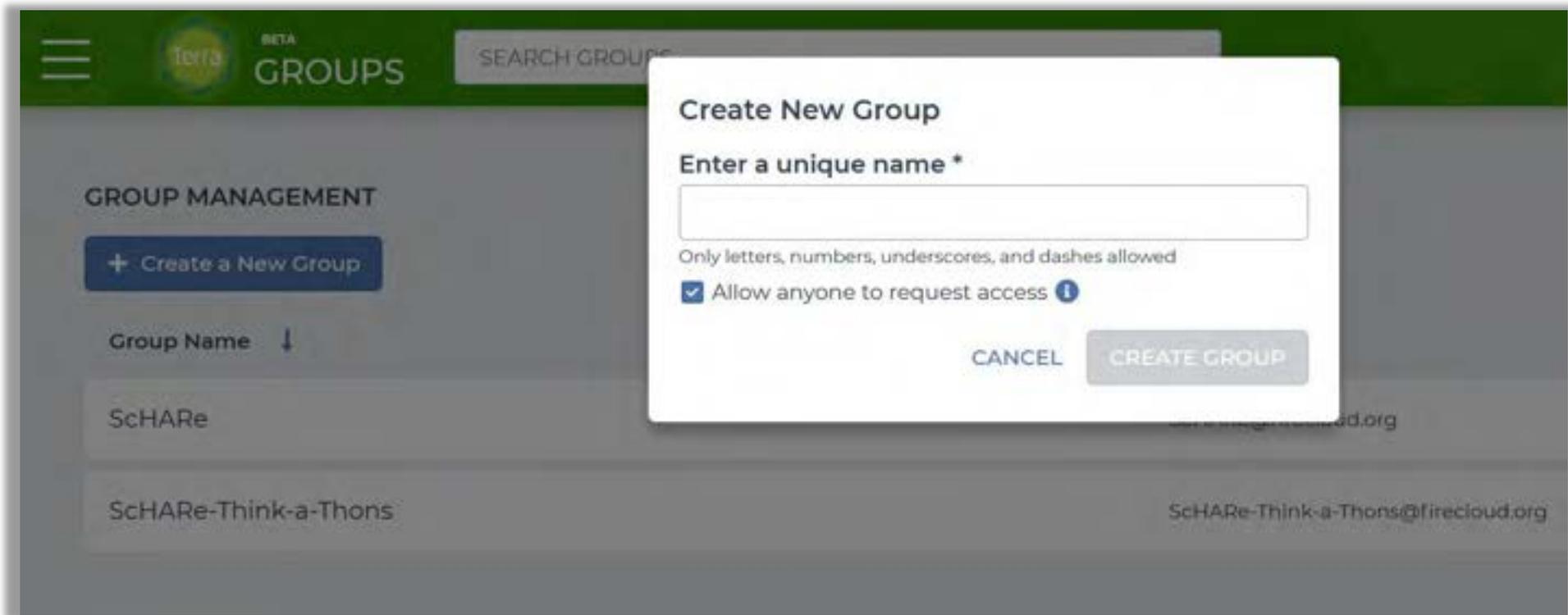
Collaborators and permissions

Click on the menu in the top left corner of the page, then on “Groups”



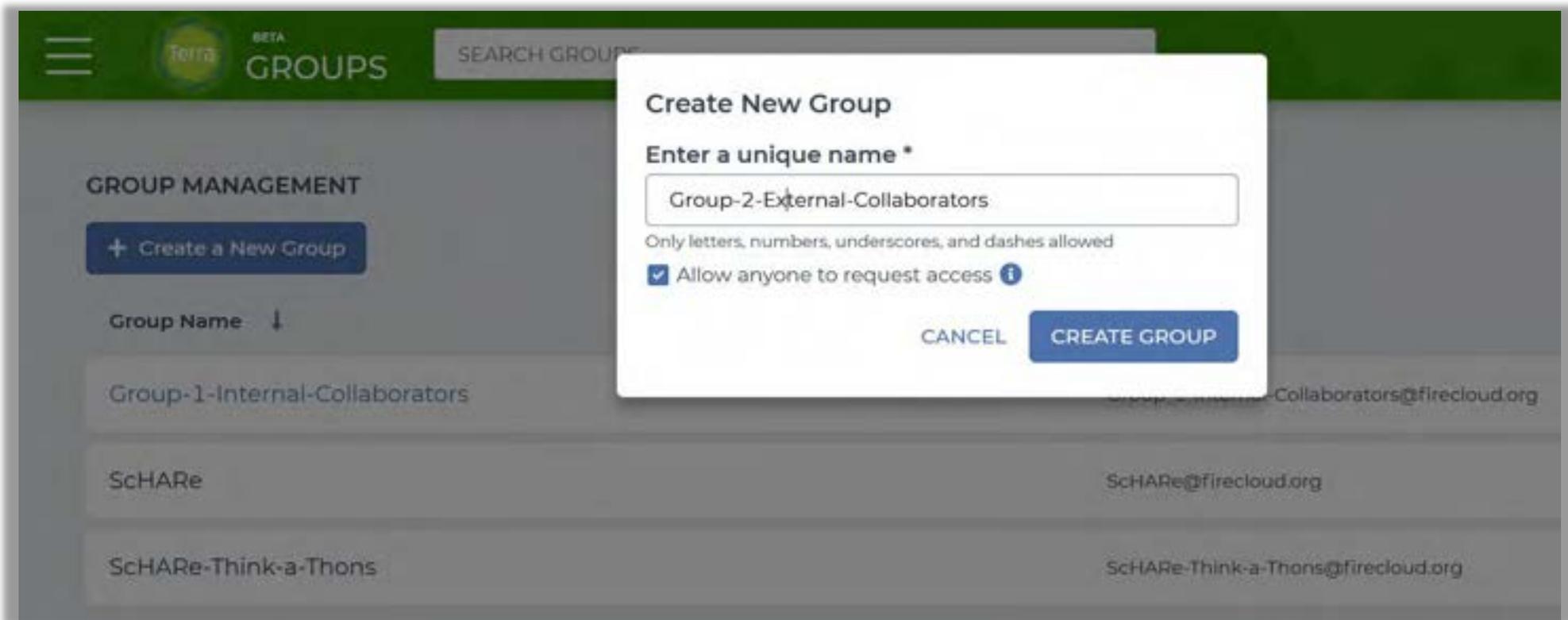
Collaborators and permissions

On the Groups page, select “Create a New Group” and proceed to **create two different groups**, one for each of the two groups of collaborators previously identified



Collaborators and permissions

On the Groups page, select “Create a New Group” and proceed to **create two different groups**, one for each of the two groups of collaborators previously identified



Collaborators and permissions

For each group, click on the name of the group and, in the following screen, on “**Add User**”



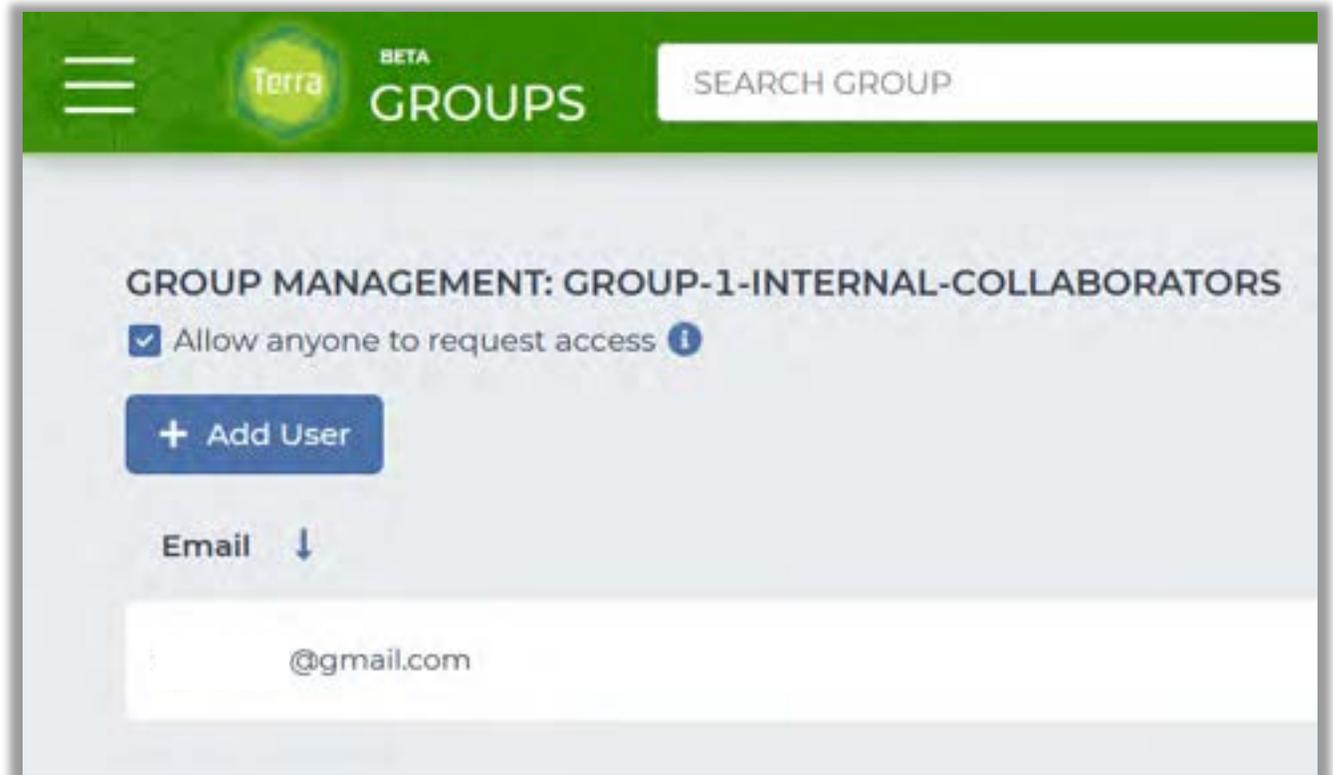
GROUP MANAGEMENT

+ Create a New Group

Group Name ↓

Group-1-Internal-Collaborators

Group-2-External-Collaborators



GROUP MANAGEMENT: GROUP-1-INTERNAL-COLLABORATORS

Allow anyone to request access ⓘ

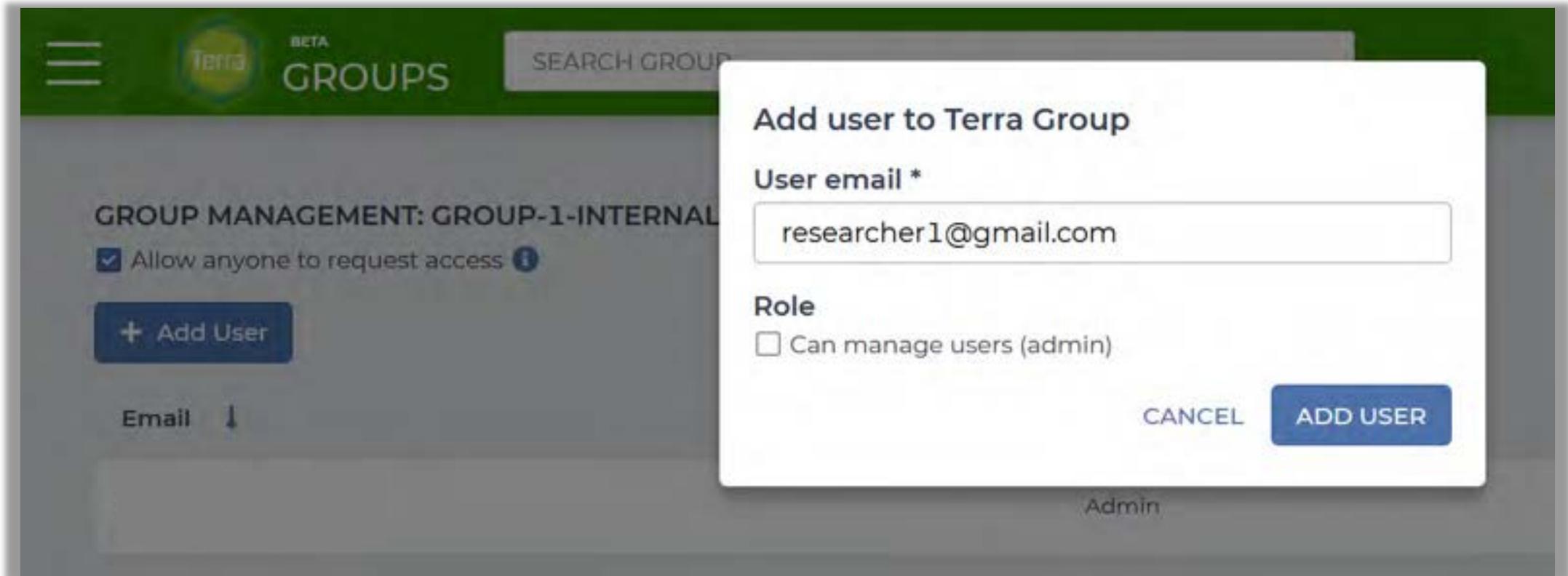
+ Add User

Email ↓

@gmail.com

Collaborators and permissions

Add the **Google email address of at least one researcher** to each group. If you want one or more of your collaborators to be able to manage users and groups, check the “Can manage users (admin)” box



Authorization domain

You now have two lists of collaborators with whom you can share your workspace, assigning different roles. It's time to create an **authorization domain** for your workspace!

What is an authorization domain (AD)?

Data in the cloud is more secure than data stored locally. Terra has several layers of protection.

ADs are data protection mechanisms - **managed groups with strictly enforced workspace permissions** that, like a badge, follow the original workspace and all copies of it.

The AD guarantees that only members of the AD can clone (copy) the original workspace. If anyone tries to share a cloned (copied) workspace with a colleague who doesn't have the right badge (doesn't belong to the AD), the recipient won't be able to view the workspace. **Your data and analyses are safe!**

Authorization domain

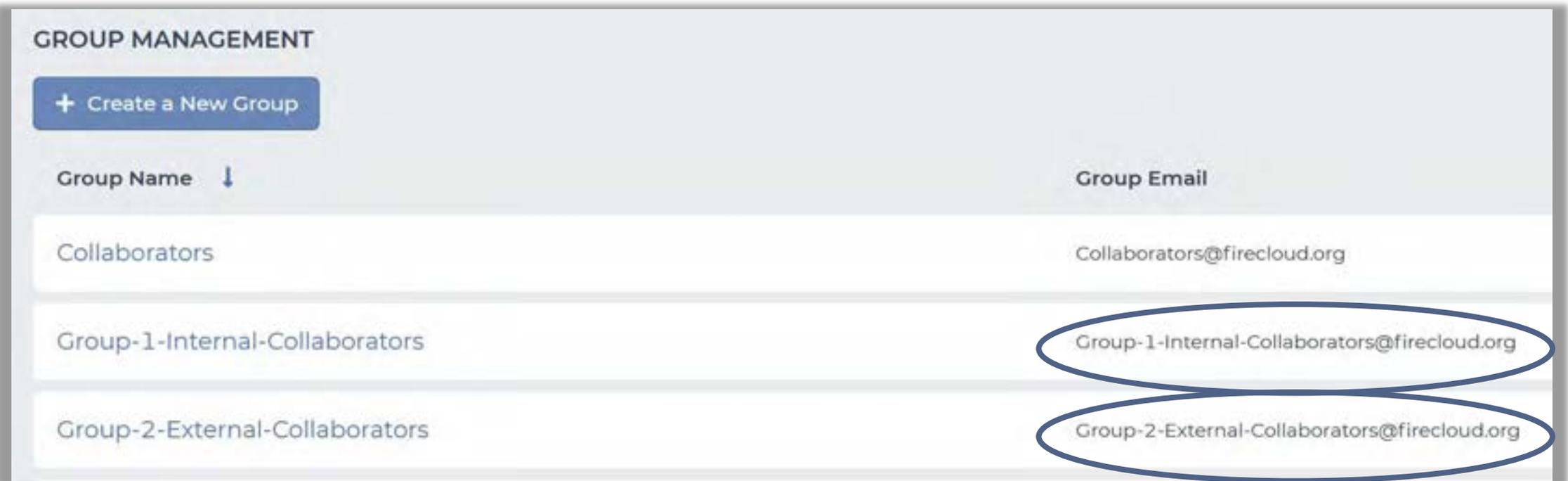
Create an **authorization domain** for your workspace, by adding a new group called “Collaborators” in the “Groups” section.

The image consists of three sequential screenshots illustrating the process of creating a new group in Terra:

- Left Screenshot:** Shows the 'GROUP MANAGEMENT' interface. At the top, there is a green header with the Terra logo and 'BETA GROUPS'. Below the header, there is a search bar labeled 'SEARCH GROUPS'. The main content area has a 'GROUP MANAGEMENT' section with a '+ Create a New Group' button. Below this, there is a 'Group Name' dropdown menu and a list of existing groups: 'Group-1-Internal-Collaborators', 'Group-2-External-Collaborators', 'ScHARe', and 'ScHARe-Think-a-Thons'.
- Middle Screenshot:** Shows a 'Create New Group' modal dialog. It prompts the user to 'Enter a unique name *' and has a text input field containing 'Collaborators'. Below the input field, there is a note: 'Only letters, numbers, underscores, and dashes allowed'. There is also a checked checkbox labeled 'Allow anyone to request access' with an information icon. At the bottom of the modal, there are 'CANCEL' and 'CREATE GROUP' buttons.
- Right Screenshot:** Shows the 'GROUP MANAGEMENT' interface after the new group has been created. The '+ Create a New Group' button is still present. The 'Group Name' dropdown is now set to 'Collaborators'. The list of groups now includes 'Collaborators' at the top, followed by 'Group-1-Internal-Collaborators', 'Group-2-External-Collaborators', 'ScHARe', and 'ScHARe-Think-a-Thons'.

Authorization domain

Both your groups of collaborators will need to **belong to your “Collaborators” authorization domain** to be able to access your workspace. Simply add the group email associated with each group (circled in blue below) to the “Collaborators” group list: every email address in your groups will now belong to the “Collaborators” authorization domain as well. You can add additional users to each group at any time



GROUP MANAGEMENT	
Group Name ↓	Group Email
Collaborators	Collaborators@firecloud.org
Group-1-Internal-Collaborators	Group-1-Internal-Collaborators@firecloud.org
Group-2-External-Collaborators	Group-2-External-Collaborators@firecloud.org

Authorization domain

The image shows a screenshot of the Terra Groups management interface. A modal window titled "Add user to Terra Group" is open, displaying a form to add a new user. The form includes a "User email *" field with the value "Group-1-Internal-Collaborators@firecloud.org" and a "Role" section with a checkbox for "Can manage users (admin)". The modal has "CANCEL" and "ADD USER" buttons.

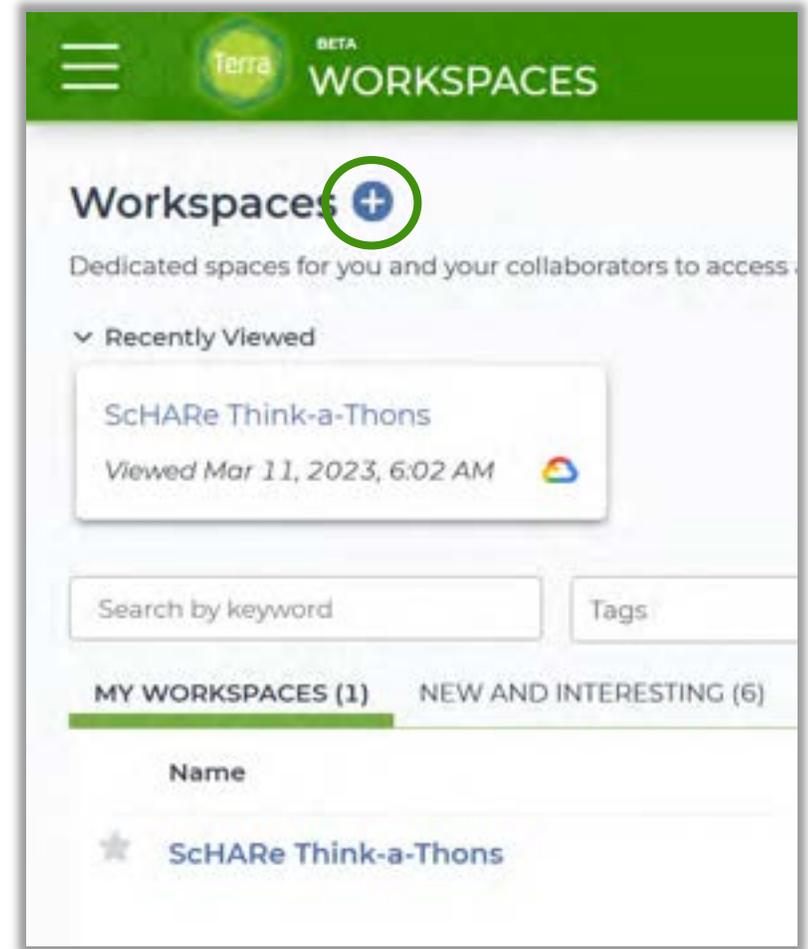
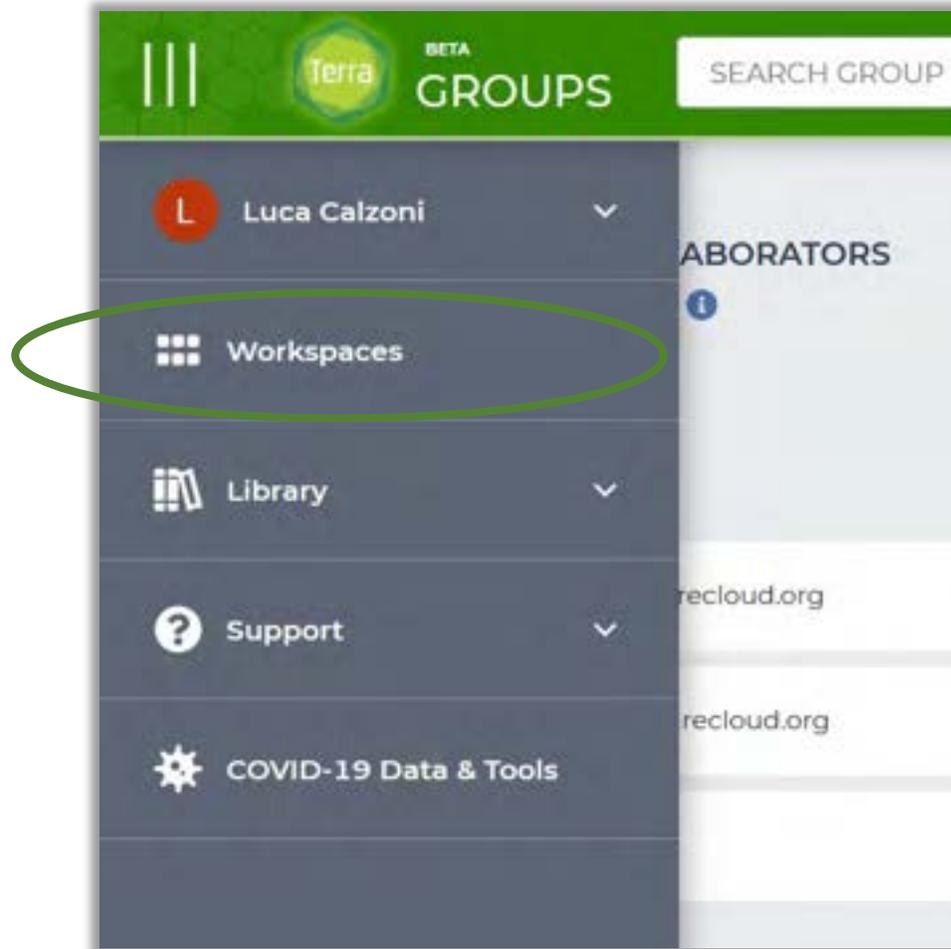
In the background, the "GROUP MANAGEMENT: COLLABORATORS" section is visible, showing a checked option for "Allow anyone to request access" and an "Add User" button. Below this, a table lists the current collaborators:

Email	Roles
Group-1-Internal-Collaborators@firecloud.org	Member
Group-2-External-Collaborators@firecloud.org	Member

Creating the workspace

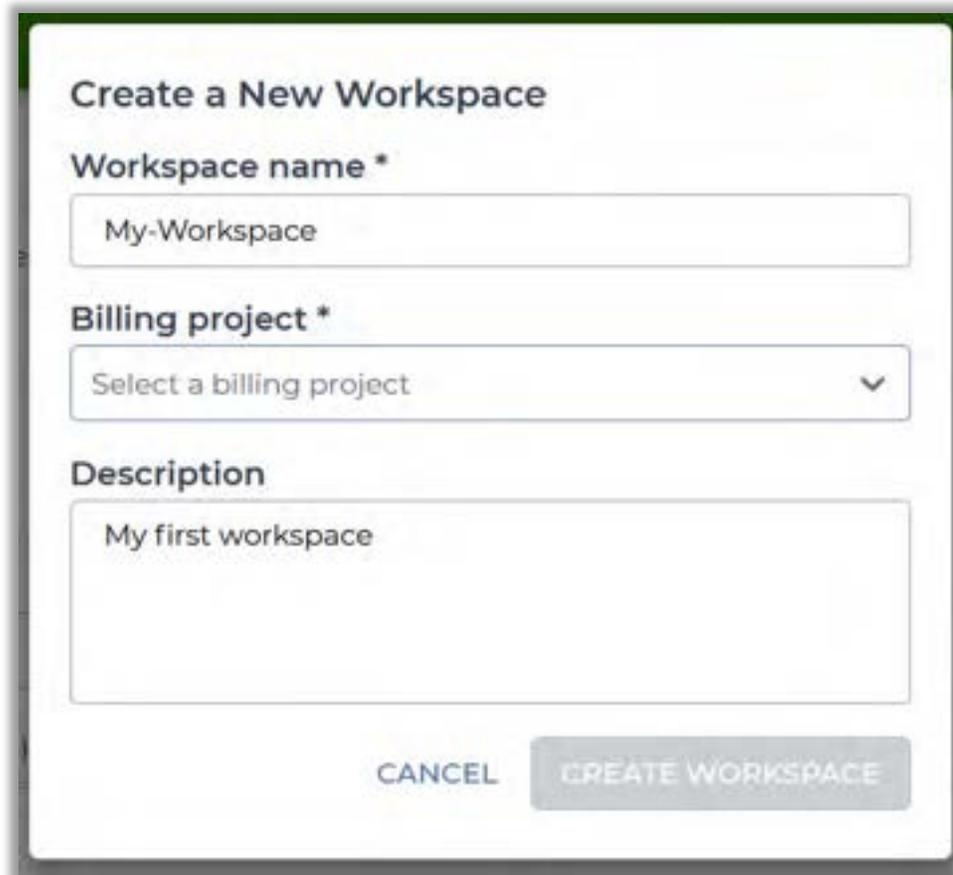
It's now time to create your workspace!

In the "Workspaces" section, click on the "+" sign next to "Workspaces"



Creating the workspace

Assign a **name** to your workspace, select our free temporary **billing project** “ScHARe-Temp” (or create one following the instructions provided during this Think-a-Thon), and add a **description**



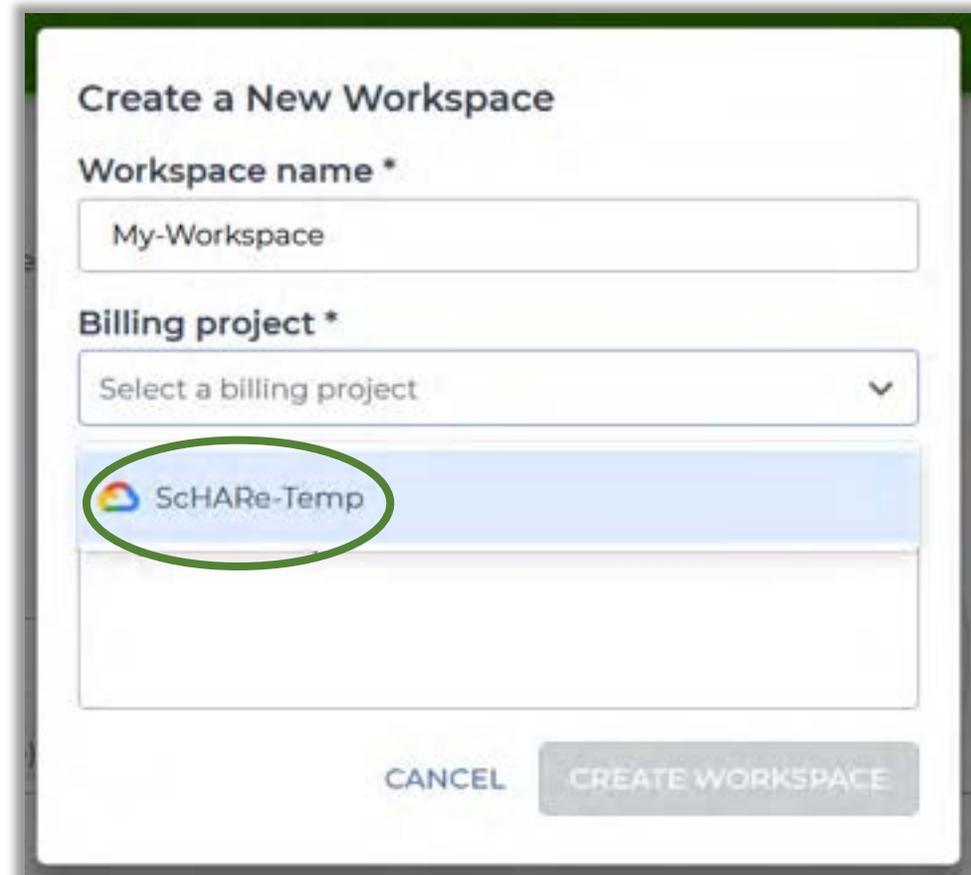
Create a New Workspace

Workspace name *

Billing project *

Description

CANCEL **CREATE WORKSPACE**



Create a New Workspace

Workspace name *

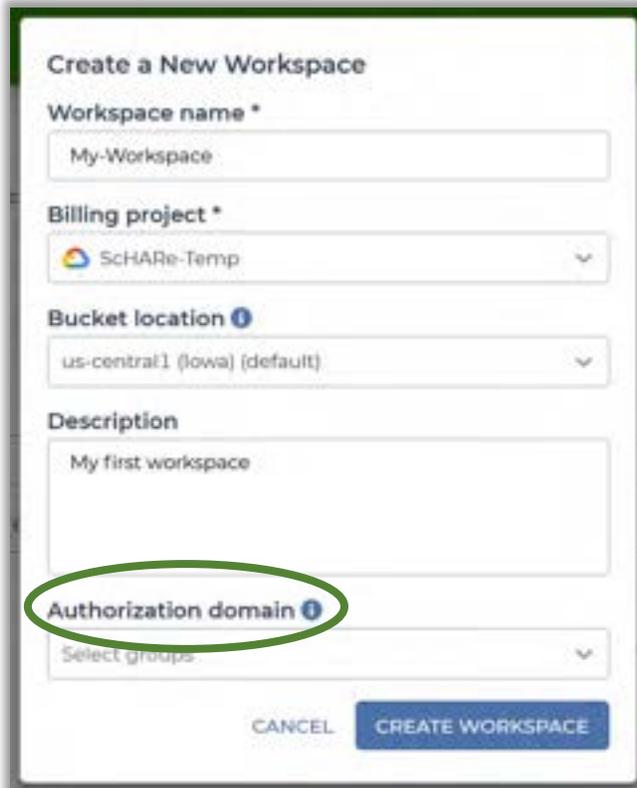
Billing project *

ScHARe-Temp

CANCEL **CREATE WORKSPACE**

Creating the workspace

Assign the “Collaborators” authorization domain to the workspace



Create a New Workspace

Workspace name *
My-Workspace

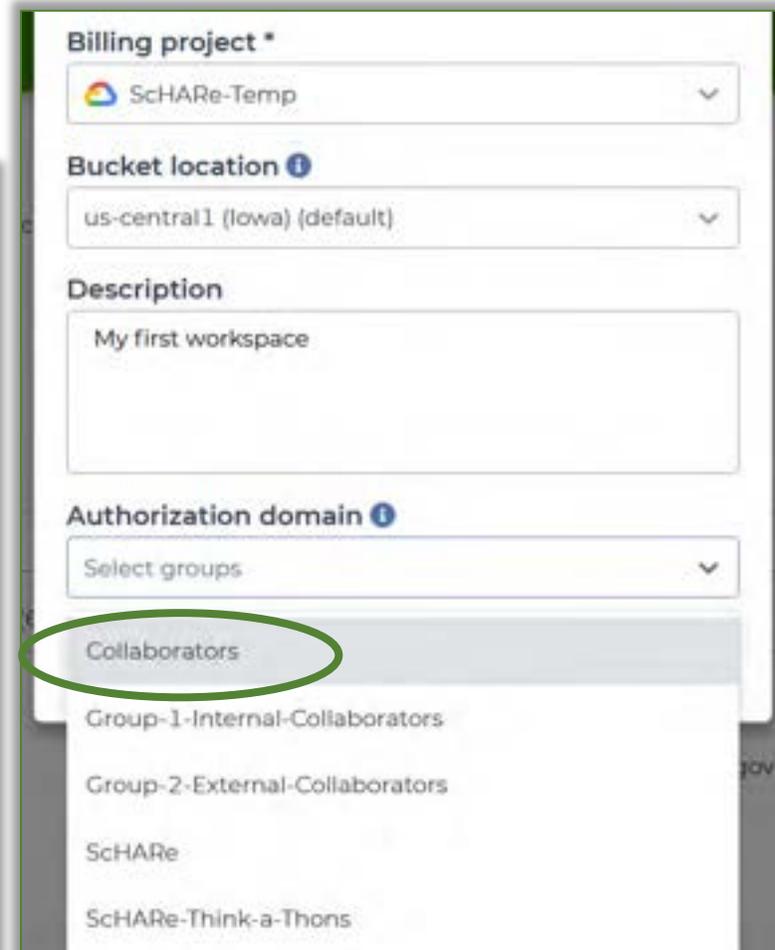
Billing project *
ScHARe-Temp

Bucket location ⓘ
us-central1 (Iowa) (default)

Description
My first workspace

Authorization domain ⓘ
Select groups

CANCEL CREATE WORKSPACE



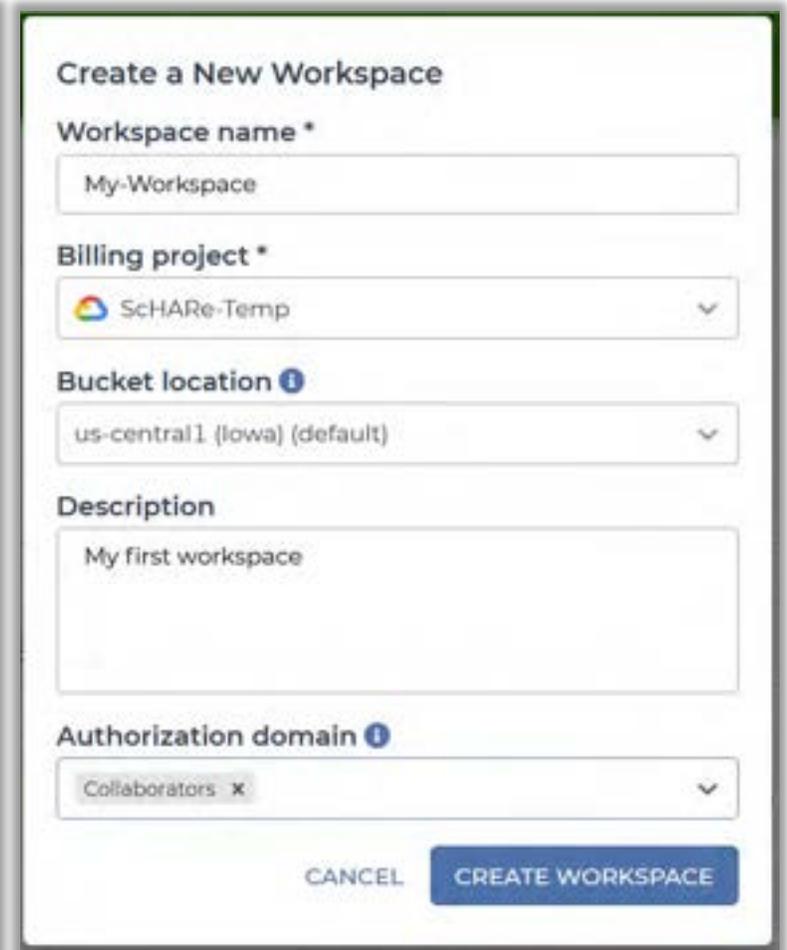
Billing project *
ScHARe-Temp

Bucket location ⓘ
us-central1 (Iowa) (default)

Description
My first workspace

Authorization domain ⓘ
Select groups

- Collaborators
- Group-1-Internal-Collaborators
- Group-2-External-Collaborators
- ScHARe
- ScHARe-Think-a-Thons



Create a New Workspace

Workspace name *
My-Workspace

Billing project *
ScHARe-Temp

Bucket location ⓘ
us-central1 (Iowa) (default)

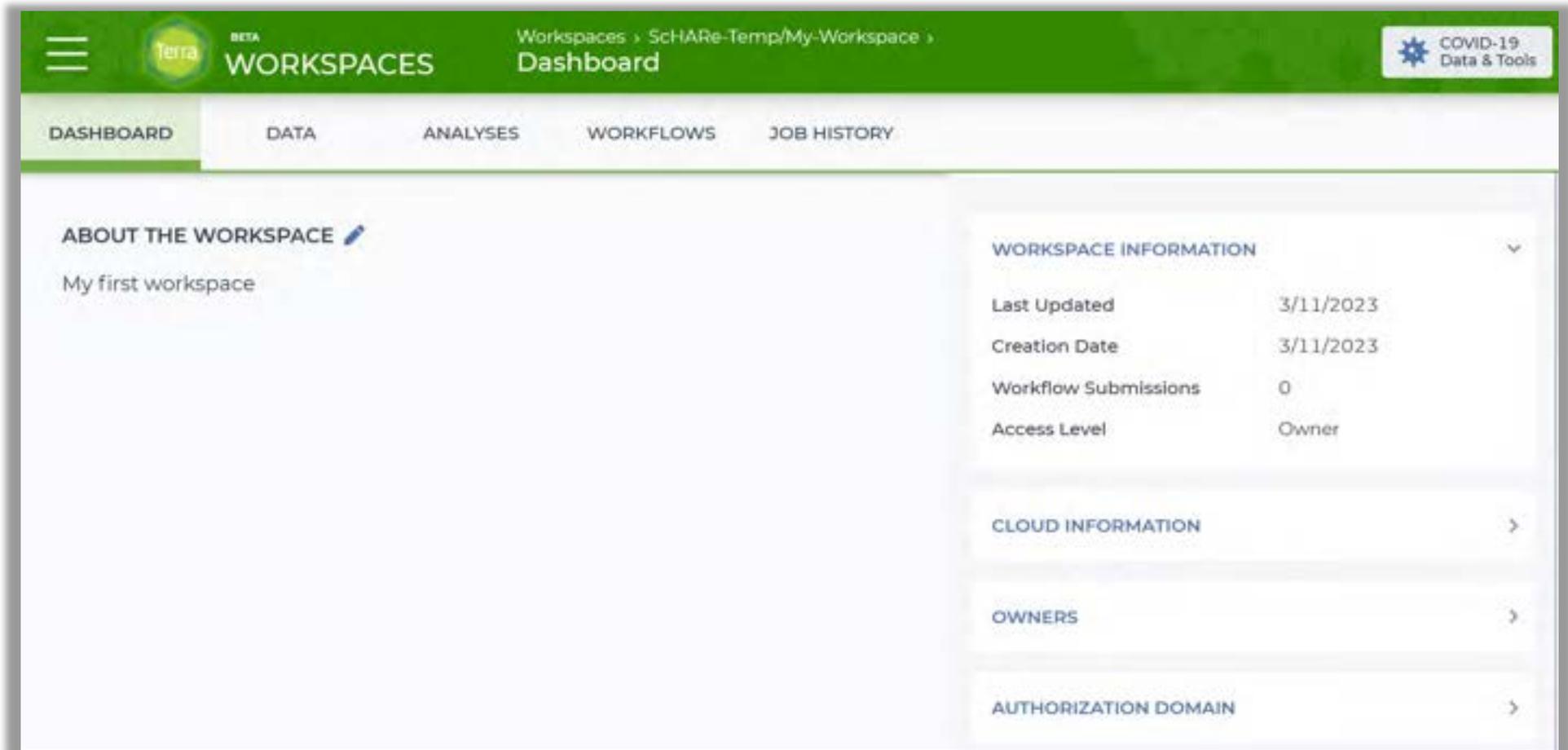
Description
My first workspace

Authorization domain ⓘ
Collaborators x

CANCEL CREATE WORKSPACE

Creating the workspace

Congratulations, your workspace was successfully created!



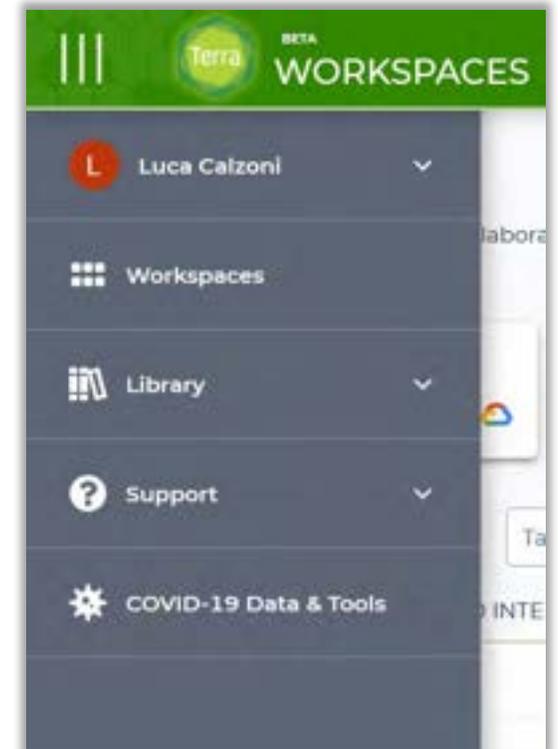
The screenshot displays the Terra WORKSPACES dashboard. The top navigation bar is green and contains the Terra logo, the word "WORKSPACES", and the breadcrumb "Workspaces > SchARe-Temp/My-Workspace > Dashboard". A "COVID-19 Data & Tools" button is visible in the top right. Below the navigation bar, there are tabs for "DASHBOARD", "DATA", "ANALYSES", "WORKFLOWS", and "JOB HISTORY". The main content area is divided into two columns. The left column has a heading "ABOUT THE WORKSPACE" with a pencil icon and the text "My first workspace". The right column contains several expandable sections: "WORKSPACE INFORMATION" (expanded), "CLOUD INFORMATION", "OWNERS", and "AUTHORIZATION DOMAIN". The "WORKSPACE INFORMATION" section contains the following data:

WORKSPACE INFORMATION	
Last Updated	3/11/2023
Creation Date	3/11/2023
Workflow Submissions	0
Access Level	Owner

Sharing the workspace

You are now ready to **share it with the two groups of collaborators** you created:

1. **Click on the menu** in the top left corner of the page, then on **“Workspaces”**
2. **Identify your workspace** in the list of workspaces provided on screen and click on the corresponding **vertical three-dot menu (A)**, then on **“Share” (B)**



Sharing the workspace

In the drop-down menu, select the group email corresponding to your **first group of internal collaborators**

Share Workspace

User email

Add people or groups ADD

Current Collaborators

@gmail.com

Owner Can share Can compute

Share with Support No CANCEL SAVE

Share Workspace

User email

Add people or groups ADD

Group-1-Internal-Collaborators@firecloud.org

Collaborators@firecloud.org

Group-2-External-Collaborators@firecloud.org

ScHARe@firecloud.org

ScHARe-Think-a-Thons@firecloud.org

Share with Support No CANCEL SAVE

Sharing the workspace

Share Workspace

User email

Add people or groups ADD

Current Collaborators

@gmail.com
Owner Can share Can compute

Group-1-Internal-Collaborators@firecloud.org
Writer Can share Can compute

Reader

Writer

Owner

CANCEL SAVE

Assigning permissions – Group 1 (Writer)

Since you want this group to be able to:

- access your data
 - perform computations
 - help you to write the notebooks you will use to share results with the scientific community
- select the “**Writer**” role for this group in the drop-down menu and check the “**Can compute**” box

Any user belonging to the group will be able to **modify** your workspace and analysis notebooks, **and perform computations** for which you will be billed through your Terra Billing Project

Notice how you are the owner of the workspace, and can add other co-owners if desired

Sharing the workspace

Share Workspace

User email
Add people or groups ADD

Current Collaborators

@gmail.com	Owner	<input checked="" type="checkbox"/> Can share	<input checked="" type="checkbox"/> Can compute
Group-1-Internal-Collaborators@firecloud.org	Writer	<input type="checkbox"/> Can share	<input checked="" type="checkbox"/> Can compute
Group-2-External-Collaborators@firecloud.org	Reader	<input checked="" type="checkbox"/> Can share	<input type="checkbox"/> Can compute

Share with Support No CANCEL SAVE

Assigning permissions – Group 2 (Reader)

Since you want this group to be able to see your data, notebooks and analyses, but without the possibility of modifying them, select the “**Reader**” role for this group in the drop-down menu and do not check the “**Can compute**” box.

If you also want the group to be able to share your work, check on the “**Can share**” box

Any user belonging to the group will be able to **see, but not modify** your workspace and analysis notebooks, and they will **not be able to perform computations** – unless they copy your notebooks and create their own Terra Billing Project, sustaining the cost of the computations

Billing permissions

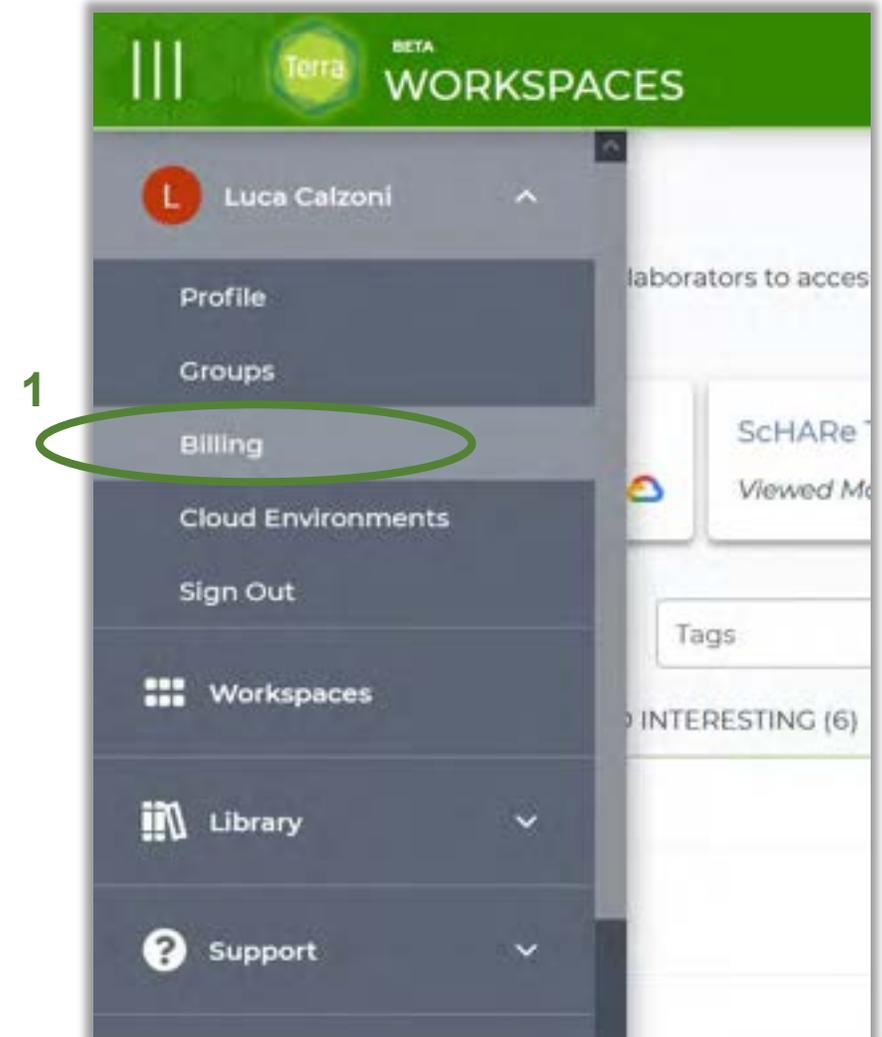
To allow collaborators from Group 1 to perform **computations** for which you will sustain the cost, you have to give them **permission to use your Terra Billing Project**

We will demonstrate how you can do this using the free temporary billing project associated to your workspace

Note - This part of the tutorial is **for demonstration purposes only**. As owners of the free billing project you used to create your demo workspace, we are the only ones able to share it with others. Revisit these slides **after creating your own billing project: you will be able to replicate these steps** using the billing project you own

Instructions:

1. **Click on the menu** in the top left corner of the page, then on **“Billing”**



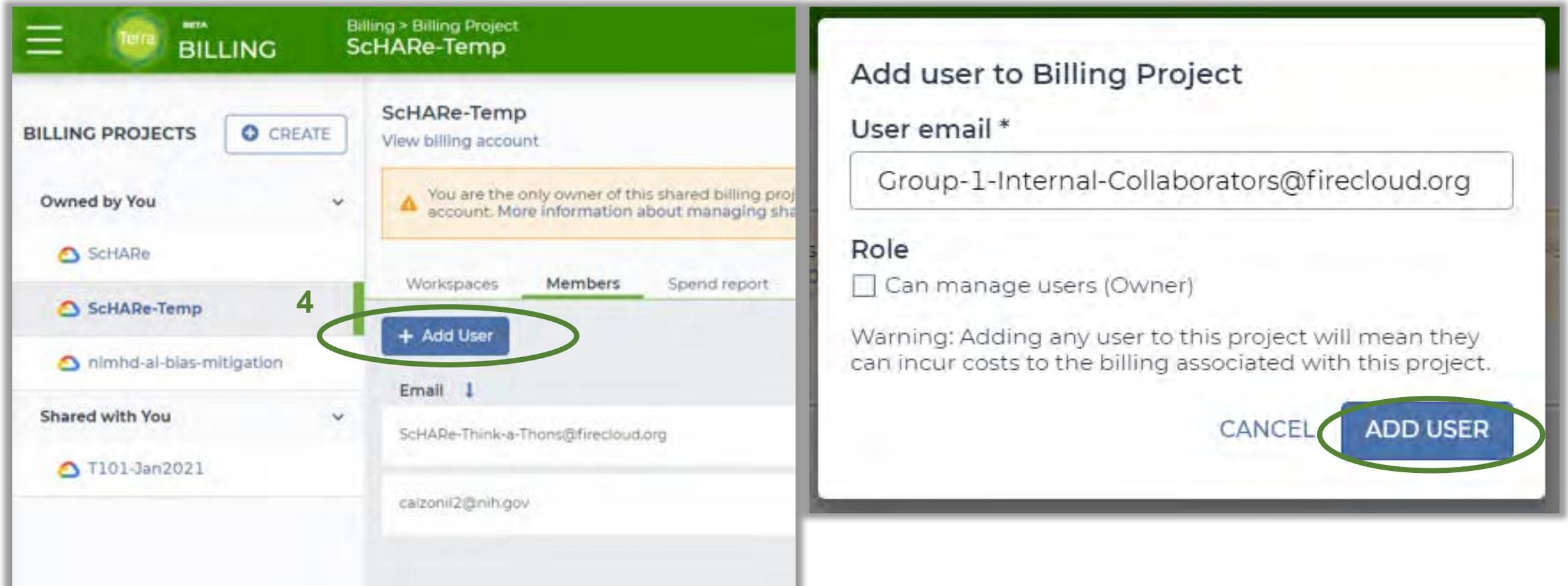
Billing permissions

2. Identify the **billing project associated with your workspace** in the list of Billing Projects owned by you, and **click on it**
3. Click on the **“Members”** tab

The screenshot displays the Terra Billing interface. The top navigation bar is green and contains the Terra logo, the word 'BETA', and the text 'BILLING'. The current page is titled 'Billing > Billing Project ScHARe-Temp'. On the left, under 'BILLING PROJECTS', there is a 'CREATE' button and a dropdown menu for 'Owned by You'. The 'ScHARe-Temp' project is highlighted with a green circle and a '2' next to it. Below it are 'nimhd-ai-bias-mitigation' and 'T101-Jan2021' under the 'Shared with You' section. On the right, the 'ScHARe-Temp' project details are shown, including a warning message: 'You are the only owner of this shared billing project account. More information about managing shared billing projects.' Below this, there are three tabs: 'Workspaces', 'Members' (circled in green with a '3' next to it), and 'Spend report'. The 'Members' tab is active, showing a list of workspaces: 'My-Workspace', 'ScHARe Think-a-Thons', 'ThinkATHonTest', and 'Wiz test'.

Billing permissions

4. Click on “Add User”, add the **group email** associated with collaborators from Group 1 to the Billing Project, and confirm. If you prefer, you can add specific **individual collaborators** instead of the entire group



The image shows a screenshot of the Terra Billing interface. On the left, a list of billing projects is shown under 'Owned by You' and 'Shared with You'. The project 'ScHARe-Temp' is highlighted, and a green circle with the number '4' is around the '+ Add User' button. On the right, a modal dialog titled 'Add user to Billing Project' is open. It contains a 'User email *' field with the value 'Group-1-Internal-Collaborators@firecloud.org', a 'Role' section with a checkbox for 'Can manage users (Owner)', and a warning message. At the bottom right of the modal, the 'ADD USER' button is circled in green.

Billing > Billing Project ScHARe-Temp

BILLING PROJECTS [CREATE](#)

Owned by You

- ScHARe
- ScHARe-Temp** 4
- nimhd-al-bias-mitigation

Shared with You

- T101-Jan2021

ScHARe-Temp
View billing account

Warning: You are the only owner of this shared billing project account. More information about managing shared accounts.

Workspaces **Members** Spend report

+ Add User

Add user to Billing Project

User email *
Group-1-Internal-Collaborators@firecloud.org

Role
 Can manage users (Owner)

Warning: Adding any user to this project will mean they can incur costs to the billing associated with this project.

[CANCEL](#) **ADD USER**

Billing permissions

Success! Your collaborators can now use your Billing Project to perform computations in the cloud

The screenshot shows the Terra Billing interface. The top navigation bar is green and contains the Terra logo, the word 'BILLING', and the breadcrumb 'Billing > Billing Project: SchARe-Temp'. On the left, a sidebar titled 'BILLING PROJECTS' has a 'CREATE' button and two sections: 'Owned by You' (containing 'SchARe' and 'SchARe-Temp') and 'Shared with You' (containing 'T101-Jan2021'). The 'SchARe-Temp' project is selected. The main content area shows the project name 'SchARe-Temp' and a 'View billing account' link. A warning message states: 'You are the only owner of this shared billing project. Consider adding another owner to ensure someone is able to manage the account. More information about managing shared billing projects.' Below this, there are tabs for 'Workspaces', 'Members', and 'Spend report'. The 'Members' tab is active, showing a '+ Add User' button and a table of members.

Email	Roles
Group-1-Internal-Collaborators@firecloud.org	User
SchARe-Think-a-Thons@firecloud.org	User
calzonil2@nih.gov	Owner

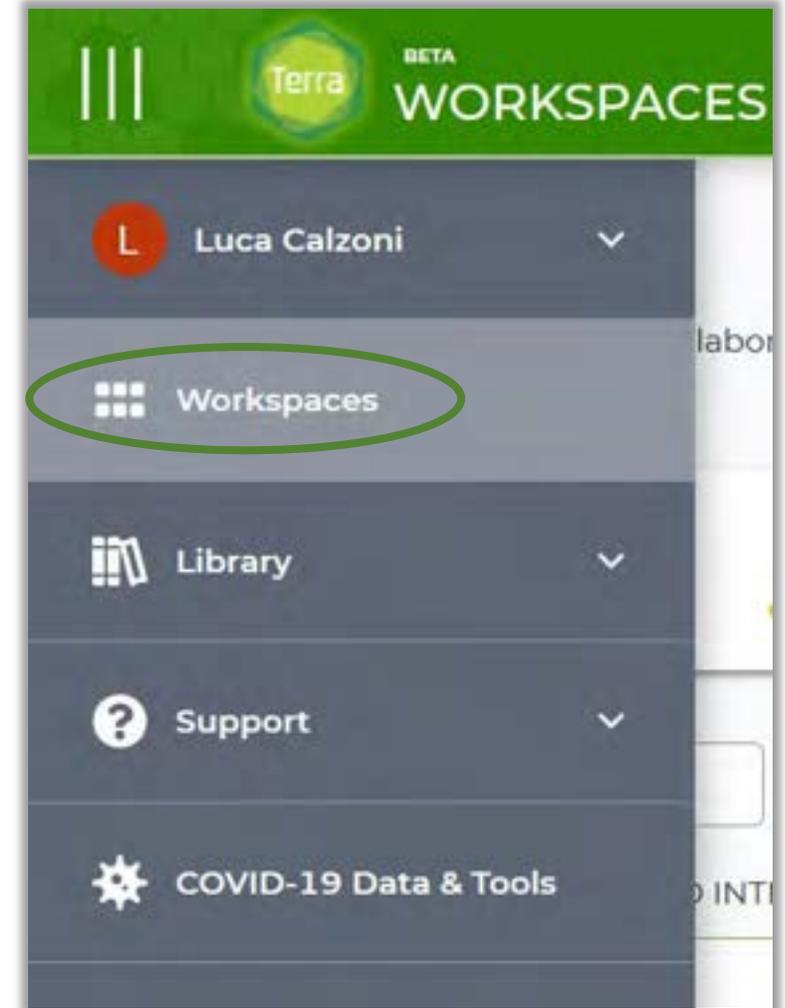
Cloning an existing workspace

One last thing. If you are interested in using the data resources of a workspace or replicating the analyses showcased in its notebooks, and have the appropriate permissions to do so, you can create a copy of such workspace for your personal use. This operation is called “cloning” the workspace

You are encouraged to clone the **ScHARe workspace** and use its resources

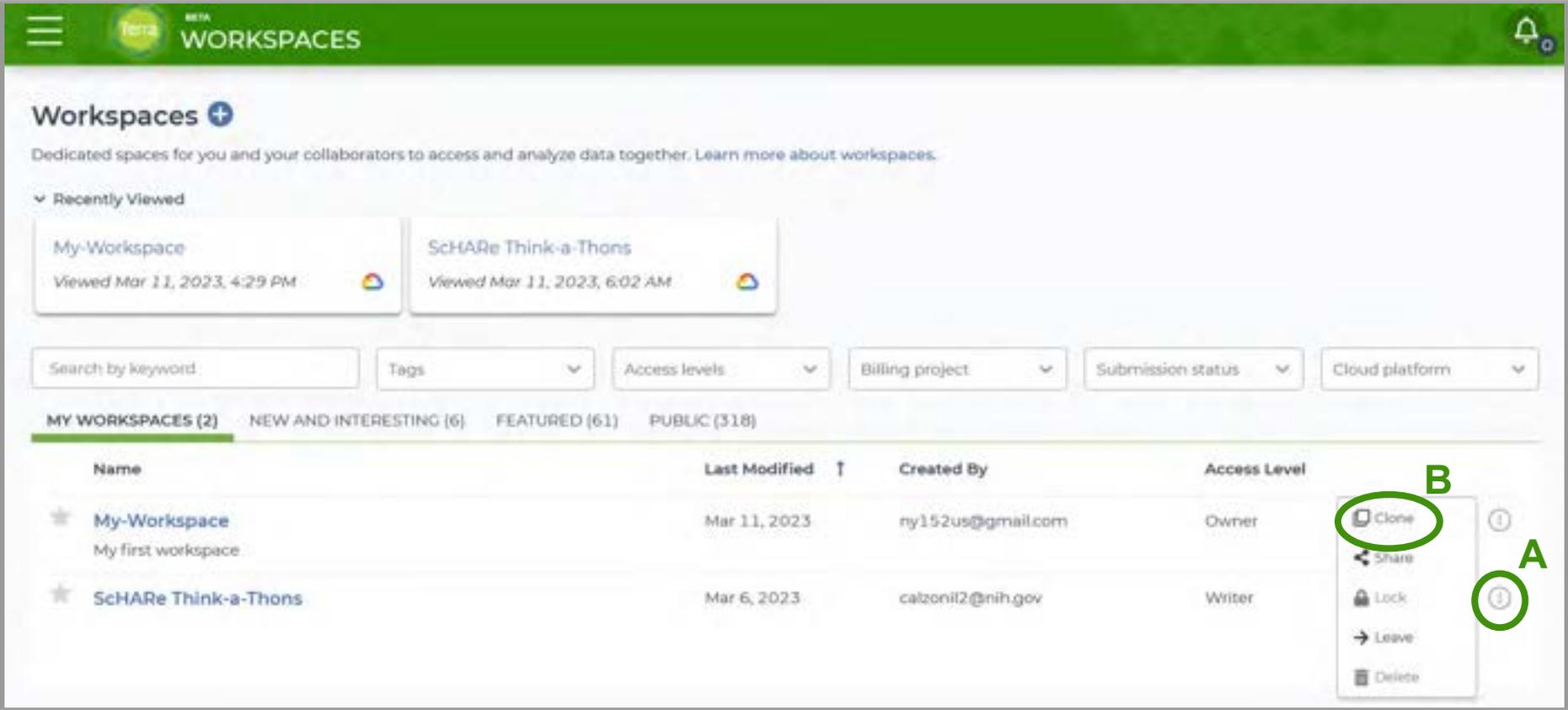
Here is how you can do it. In this example, we will clone the workspace “ScHARe Think-a-Thons”, a copy of the ScHARe workspace we created for this Think-a-Thon

1. **Click on the menu** in the top left corner of the page, then on “**Workspaces**”



Cloning an existing workspace

2. Identify the workspace you want to clone in the list of workspaces provided on screen and click on the corresponding vertical three-dot menu (A), then on “Clone” (B)



The screenshot shows the Tera Workspaces interface. At the top, there is a green header with the Tera logo and the word 'WORKSPACES'. Below the header, there is a section titled 'Workspaces' with a plus icon. Underneath, there is a 'Recently Viewed' section showing two workspace cards: 'My-Workspace' and 'SchARe Think-a-Thons'. Below this, there are several filter buttons: 'Search by keyword', 'Tags', 'Access levels', 'Billing project', 'Submission status', and 'Cloud platform'. A navigation bar shows 'MY WORKSPACES (2)' selected, along with 'NEW AND INTERESTING (6)', 'FEATURED (61)', and 'PUBLIC (318)'. The main content is a table of workspaces:

Name	Last Modified	Created By	Access Level	
My-Workspace My first workspace	Mar 11, 2023	ny152us@gmail.com	Owner	
SchARe Think-a-Thons	Mar 6, 2023	calzonil2@nih.gov	Writer	

Green circles labeled 'A' and 'B' highlight the three-dot menu icon and the 'Clone' button, respectively.

Cloning an existing workspace

3. Input a **name** for the workspace copy
4. Select the **Billing Project** you want to associate with the workspace. For this example, you can select our free temporary Billing Project “ScHARe-Temp”
5. Select the **bucket location**. A bucket location can only be set when creating a workspace. For this example, you can leave the default unmodified
6. Change the **Description** if desired
7. A cloned workspace will inherit the **Authorization Domain (AD)** groups of the original workspace. Add an additional AD if needed

Clone this workspace

Workspace name *
ScHARe Think-a-Thons copy

Billing project *
ScHARe-Temp

Bucket location ⓘ
us-central1 (Iowa) (default)

Description

Science Collaborative for Health disparities

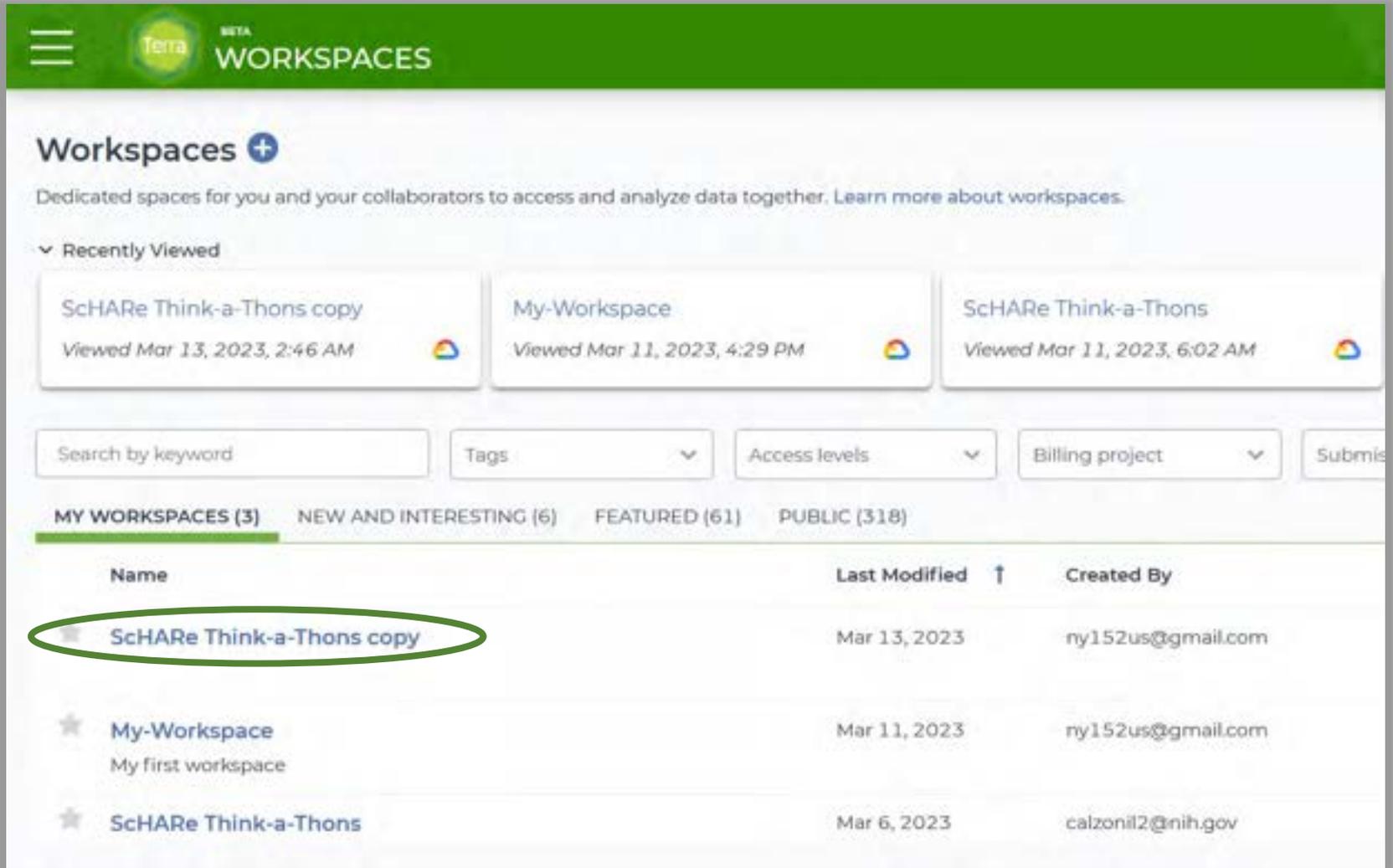
Authorization domain ⓘ
Inherited groups:
ScHARe, ScHARe-Think-a-Thons
Select groups

CANCEL CLONE WORKSPACE

Cloning an existing workspace

Success! The cloned workspace is now listed among your workspaces

You can freely access all of its resources



The screenshot shows the Terra WORKSPACES interface. At the top, there is a green header with the Terra logo and the text "BETA WORKSPACES". Below the header, the main content area is titled "Workspaces +" and includes a description: "Dedicated spaces for you and your collaborators to access and analyze data together. Learn more about workspaces." Underneath, there is a "Recently Viewed" section with three workspace cards: "ScHARe Think-a-Thons copy" (viewed Mar 13, 2023, 2:46 AM), "My-Workspace" (viewed Mar 11, 2023, 4:29 PM), and "ScHARe Think-a-Thons" (viewed Mar 11, 2023, 6:02 AM). Below this, there are search and filter options: "Search by keyword", "Tags", "Access levels", "Billing project", and "Submit". A navigation bar shows "MY WORKSPACES (3)" selected, along with "NEW AND INTERESTING (6)", "FEATURED (61)", and "PUBLIC (318)". The main workspace list has columns for "Name", "Last Modified", and "Created By". The first row, "ScHARe Think-a-Thons copy", is circled in green. The second row is "My-Workspace" (My first workspace) and the third is "ScHARe Think-a-Thons".

Name	Last Modified	Created By
ScHARe Think-a-Thons copy	Mar 13, 2023	ny152us@gmail.com
My-Workspace My first workspace	Mar 11, 2023	ny152us@gmail.com
ScHARe Think-a-Thons	Mar 6, 2023	calzonil2@nih.gov



SciARe

Part IV

Notebooks and Environment

What is a notebook?

A Jupyter Notebook is an interactive analysis tool that includes:

- **code cells** for manipulating and visualizing data in real time (Terra notebooks support **Python or R**)
- **documentation** to make it easier to share and reproduce your analysis

To get the most out of this tutorial you should be familiar with **programming**. If you are not, the code in our notebooks is very easy to understand and reuse, and our tutorial will still help you understand how notebooks work

We will:

- cover the basics of **creating your first notebook**
- **explore the instructional notebooks** available in the SchARe workspace and run one of them

Why use notebooks?

A notebook integrates code and its output into a single document where you can run code, display the output, and also add explanations, formulas, and charts

Using notebooks:

- **is now a major part of the data science workflow** at research institutions across the globe
- can make your work **more transparent, understandable, repeatable, and shareable**
- will **speed up your workflow** and make it easier to communicate and share your results

ScHARe notebooks

Let's take a first look at what a notebook can do by checking out the instructional notebooks that **ScHARe offers to help novice users** learn how to use the workspace and its resources

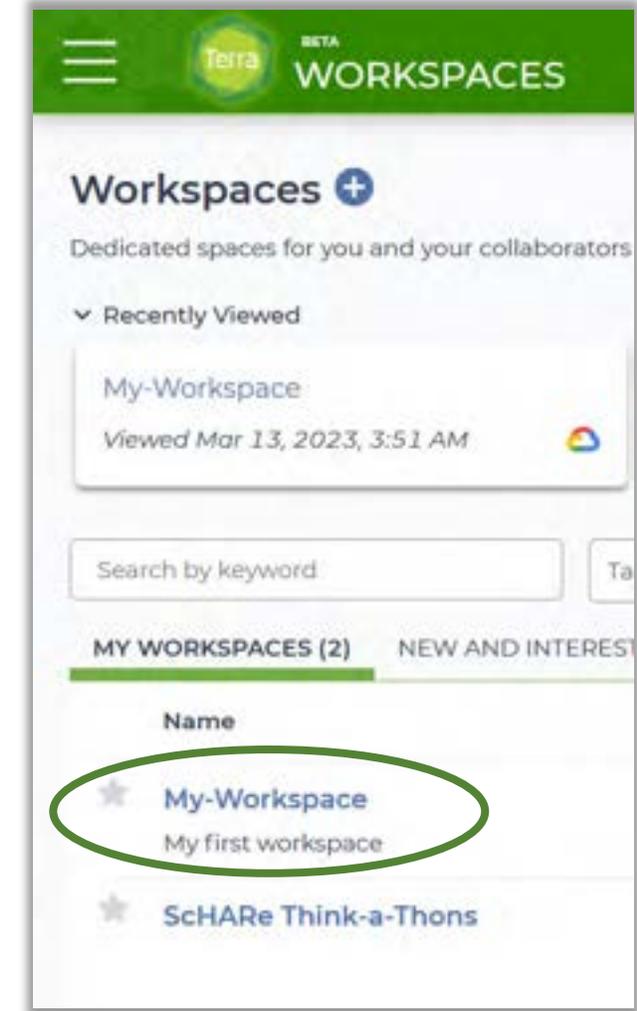
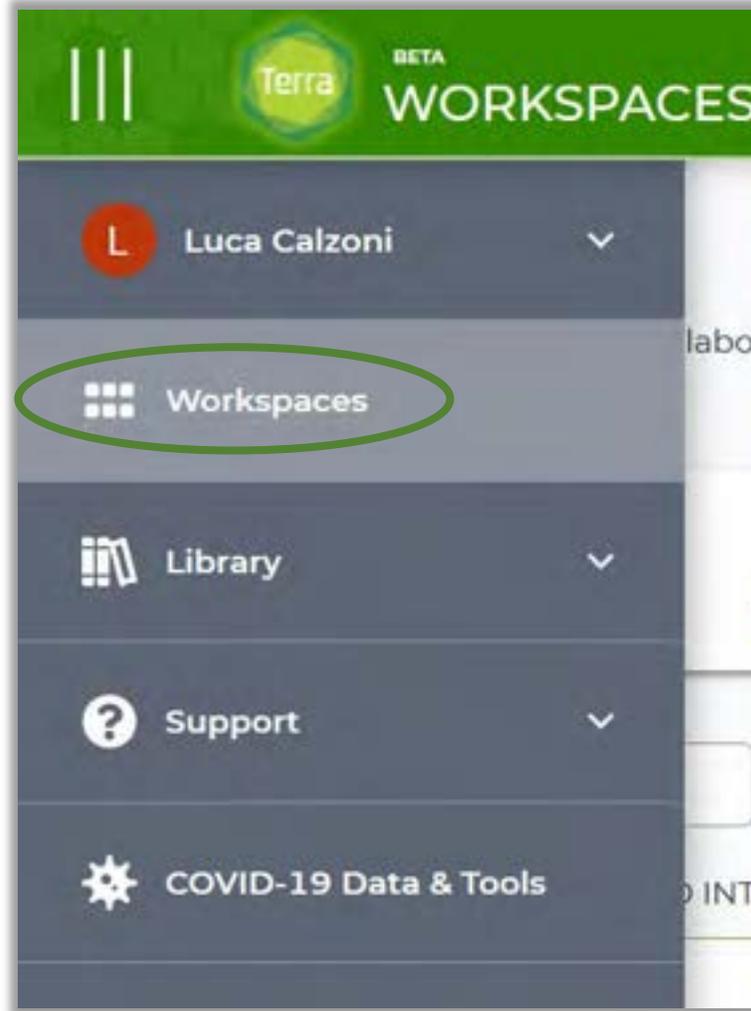
A list of the available notebooks is provided on the right. **We will access and run a copy of one of these notebooks**, as an example

List of ScHARe instructional notebooks

- **00_List of Datasets Available on ScHARe:** a list of the datasets available in the ScHARe Datasets collection.
- **01_Introduction to Terra Cloud Environment:** an introduction to the Terra platform and cloud environment.
- **02_Introduction to Terra Jupyter Notebooks:** an introduction to Jupyter Notebooks on the Terra platform.
- **03_R Environment setup:** instructions on how to setup your cloud environment for R-based notebooks.
- **04_Python 3 Environment setup:** instructions on how to setup your cloud environment for Python 3-based notebooks.
- **05_How to access plot and save data from public BigQuery datasets using R:** instructions on how to access, plot, and save data from datasets available on the cloud through the Google Cloud Public Datasets Program, using R.
- **06_How to access plot and save data from public BigQuery datasets using Python 3:** instructions on how to access, plot, and save data from datasets available on the cloud through the Google Cloud Public Datasets Program, using Python 3.
- **07_How to access plot and save data from ScHARe hosted datasets using Python 3:** instructions on how to access, plot, and save data from datasets hosted by ScHARe in this workspace.
- **08_How to upload access plot and save data stored locally using Python 3:** instructions on how to import to Terra, access, plot, and save data from datasets stored locally on your computer.

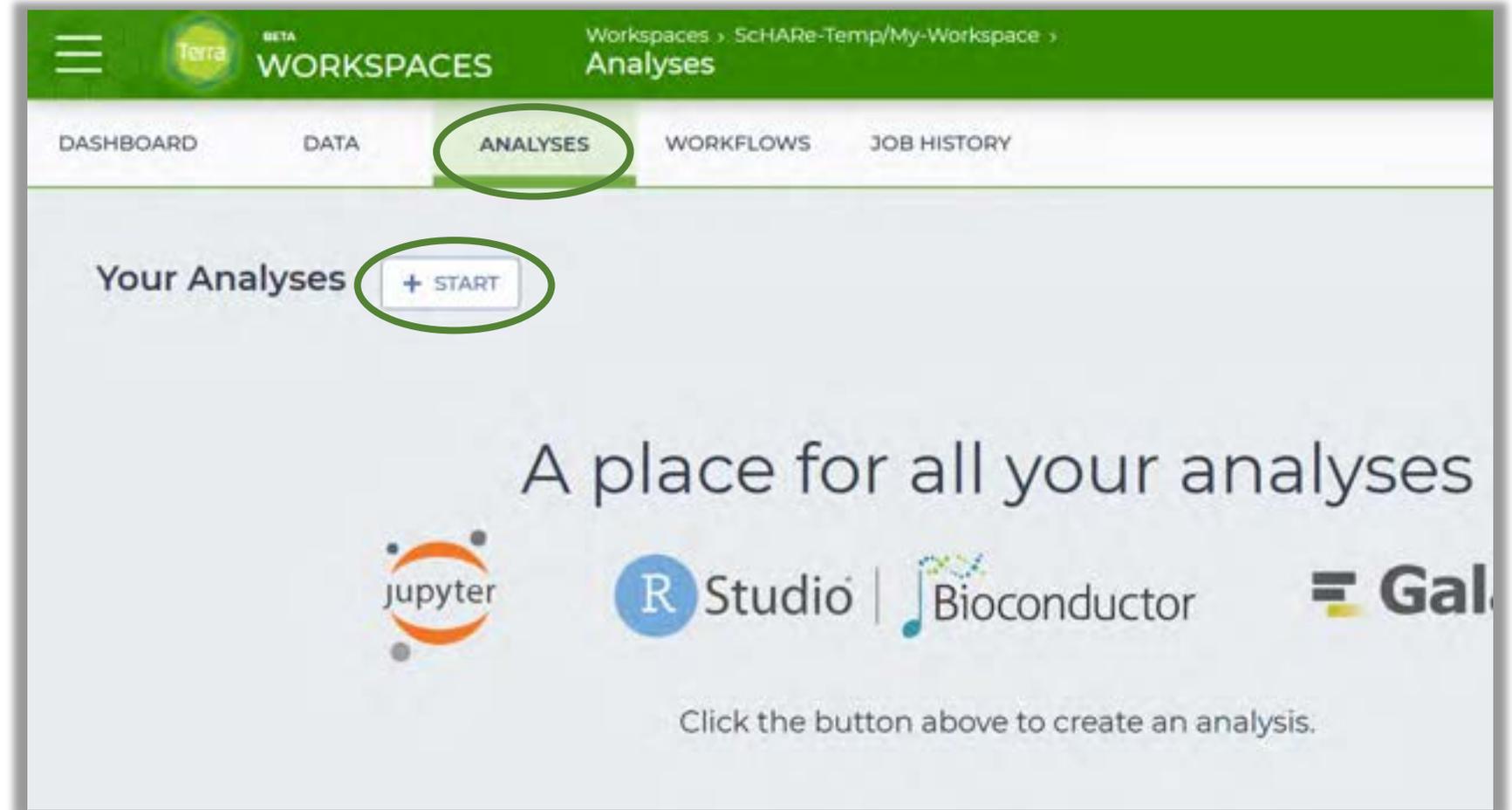
Creating a notebook

1. Click on the menu in the top left corner of the page, then on “Workspaces”
2. Click on the new workspace you created earlier today



Creating a notebook

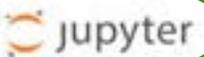
3. Click on the “Analyses” tab
4. Click on the “Start” button



Configuring the environment

5. Select “**Jupyter**”
6. In the next window, assign a **name** to the notebook and choose a programming language (“**Python 3**”)
7. Click “**Create analysis**”

Select an application

 jupyter

 Studio |  Bioconductor

 Galaxy

Or Click / Drag to upload an analysis file



Create a new notebook

Name of the notebook *

Test

Language *

Python 3

Python 3

R

Create a new notebook

Name of the notebook *

Test

Language *

Python 3

CREATE ANALYSIS

Configuring the environment

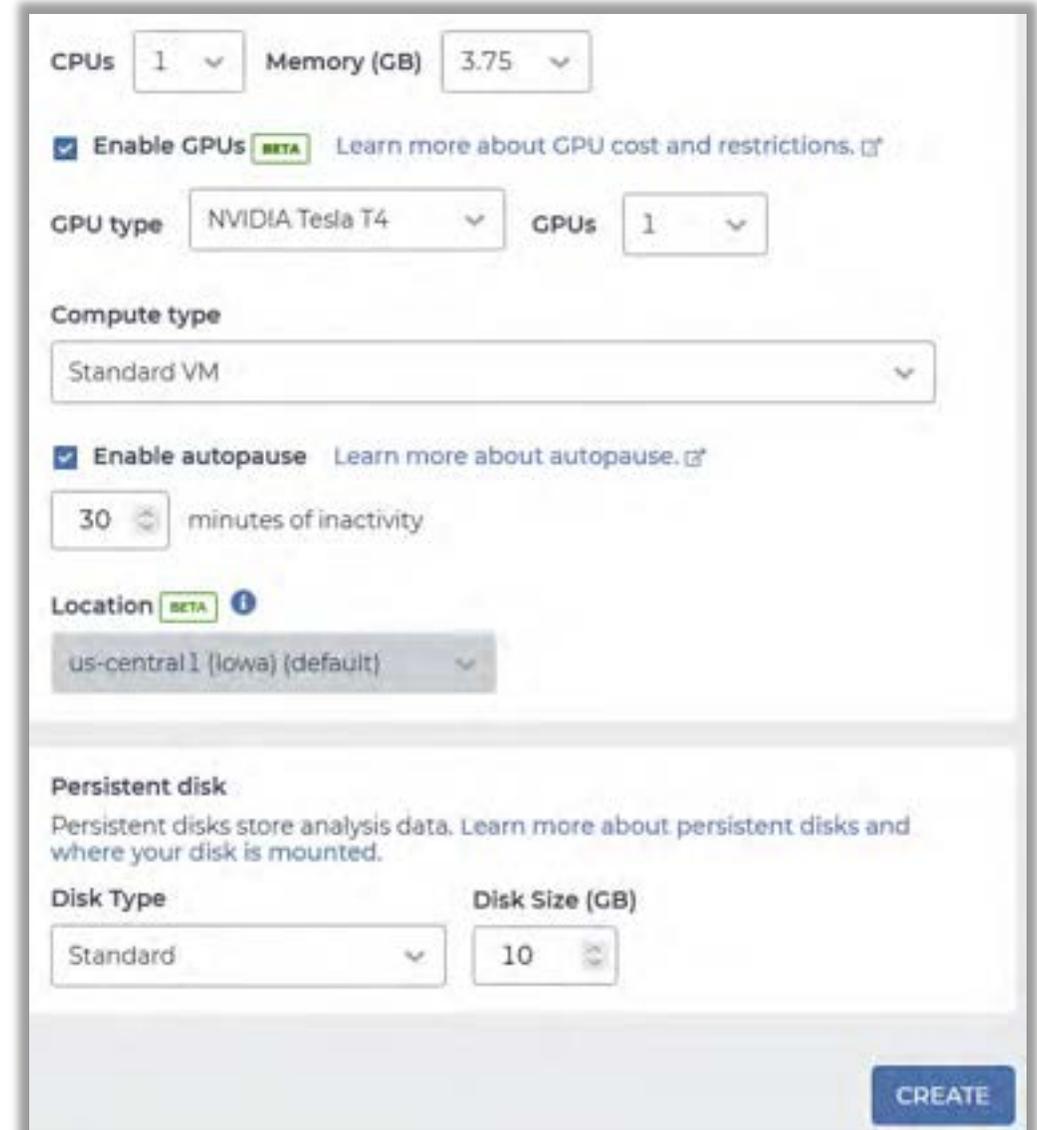
8. You will now be asked to configure your **Cloud Environment**. In this window, you can **configure** the on-demand availability of **computer system resources**, especially data storage and computing power, needed to perform your computations:
- **Application configuration:** the software application + programming languages + packages used to create your cloud environment. You can leave this unchanged
 - **Cloud compute profile:** the CPU and RAM available to run your application, which determines how much processing can be done at a time.

Notice how changing the parameters changes the estimated costs to run at the top of the window. **Find the balance that's right for you.** More compute power costs more, and you don't want to request (and pay for) significantly more than your computation needs

The screenshot shows the 'Jupyter Cloud Environment' configuration page. At the top, it displays the title 'Jupyter Cloud Environment' and a brief description: 'A cloud environment consists of application configuration, cloud compute and persistent disk(s)'. Below this, there are three cost-related metrics: 'Running cloud compute cost' at \$0.05 per hr, 'Paused cloud compute cost' at \$0.00 per hr, and 'Persistent disk cost' at \$0.40 per month. The main configuration area is divided into several sections: 'Application configuration' with a dropdown menu set to 'Default: {GATK 4.2.4.0, Python 3.7.12, R 4.2.2}', 'Startup script' with a text input field containing 'URI', 'Cloud compute profile' with 'CPUs' set to 1 and 'Memory (GB)' set to 3.75, and 'Compute type' set to 'Standard VM'. There are also checkboxes for 'Enable GPUs' (disabled) and 'Enable autopause' (checked), with a '30 minutes of inactivity' setting. The 'Location' is set to 'us-central1 (Iowa) (default)'. The interface includes various icons for help, updates, and beta features.

Configuring the environment

- **CPUs, Memory:** leave the default values unchanged
- **Enable GPUs:** checking the box allows you to add graphics processing units (GPUs) to your environment. GPUs are optimized for linear algebra computations, such as matrix multiplication, and for training artificial intelligence and deep learning models. Leave the box unchecked
- **Enable autopause:** recommended. Once your browser becomes inactive and **30** minutes pass, the cloud environment autopauses to save on costs
- **Location:** physical location where the data is going to reside. Leave the default value unchanged
- **Persistent disk:** Terra attaches a persistent disk to your cloud compute to provide an option to keep the data on the disk after you delete your compute. A **minimal cost per hour** is associated with maintaining the disk, even when the cloud compute is paused. Select “Standard” and “10 GB” and click on “**Create**”



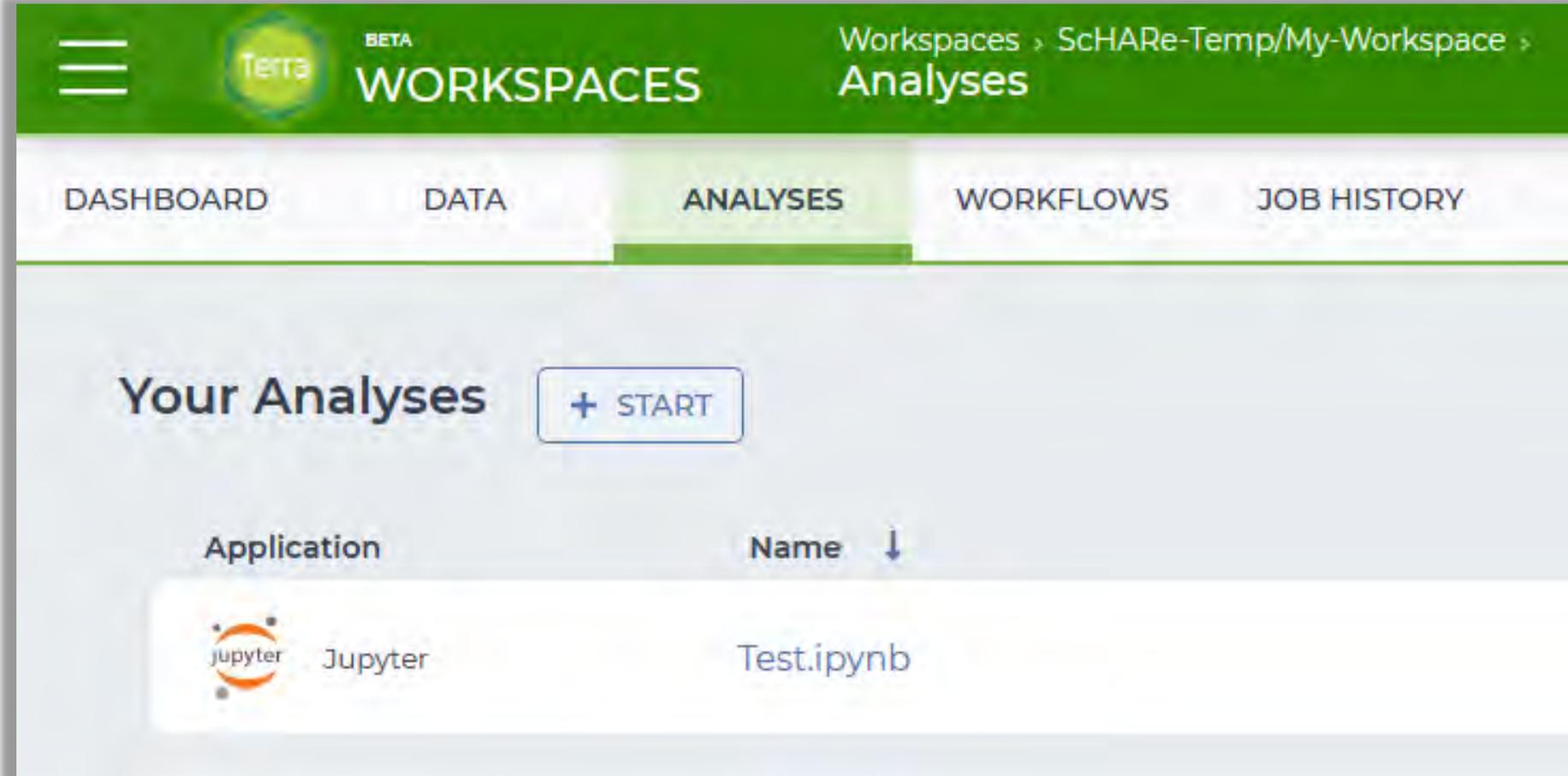
The screenshot shows the configuration page for a compute instance in Terraform Cloud. The settings are as follows:

- CPUs:** 1
- Memory (GB):** 3.75
- Enable GPUs:** (with a **BETA** label and a link to learn more about GPU cost and restrictions)
- GPU type:** NVIDIA Tesla T4
- GPUs:** 1
- Compute type:** Standard VM
- Enable autopause:** (with a link to learn more about autopause)
- Autopause duration:** 30 minutes of inactivity
- Location:** us-central1 (Iowa) (default) (with a **BETA** label and an information icon)
- Persistent disk:**
 - Disk Type:** Standard
 - Disk Size (GB):** 10

A **CREATE** button is located at the bottom right of the configuration area.

Configuring the environment

- **Success!** Your environment has been configured, and your notebook created. **Click on its name** to open it

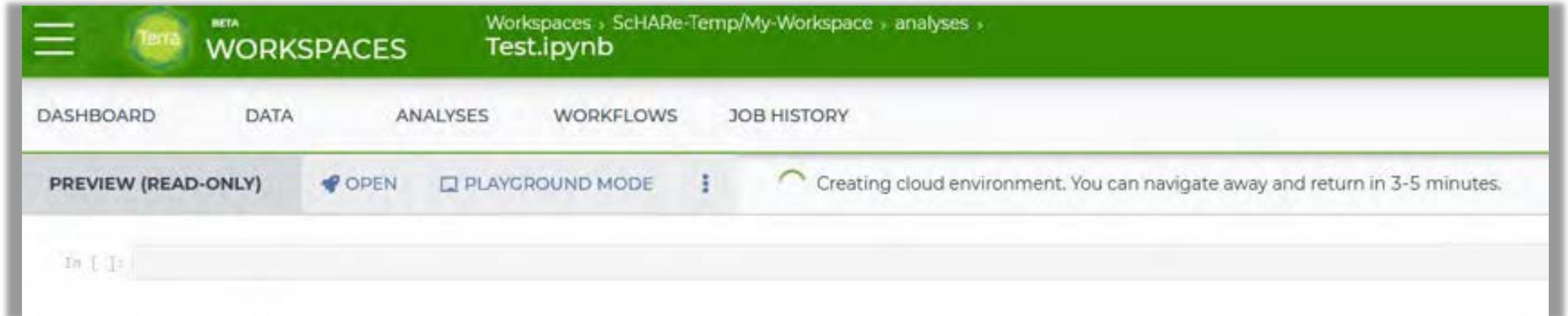


The screenshot displays the Tetta WORKSPACES interface. The top navigation bar is green and contains the Tetta logo, the word "BETA", and the text "WORKSPACES". To the right of the logo, it shows the current workspace path: "Workspaces > ScHARe-Temp/My-Workspace > Analyses". Below the navigation bar is a horizontal menu with five tabs: "DASHBOARD", "DATA", "ANALYSES" (which is highlighted with a green background), "WORKFLOWS", and "JOB HISTORY". The main content area is titled "Your Analyses" and features a "+ START" button. Below this, there is a table with two columns: "Application" and "Name". The table contains one row with the Jupyter logo in the "Application" column and "Test.ipynb" in the "Name" column.

Application	Name ↓
 Jupyter	Test.ipynb

Configuring the environment

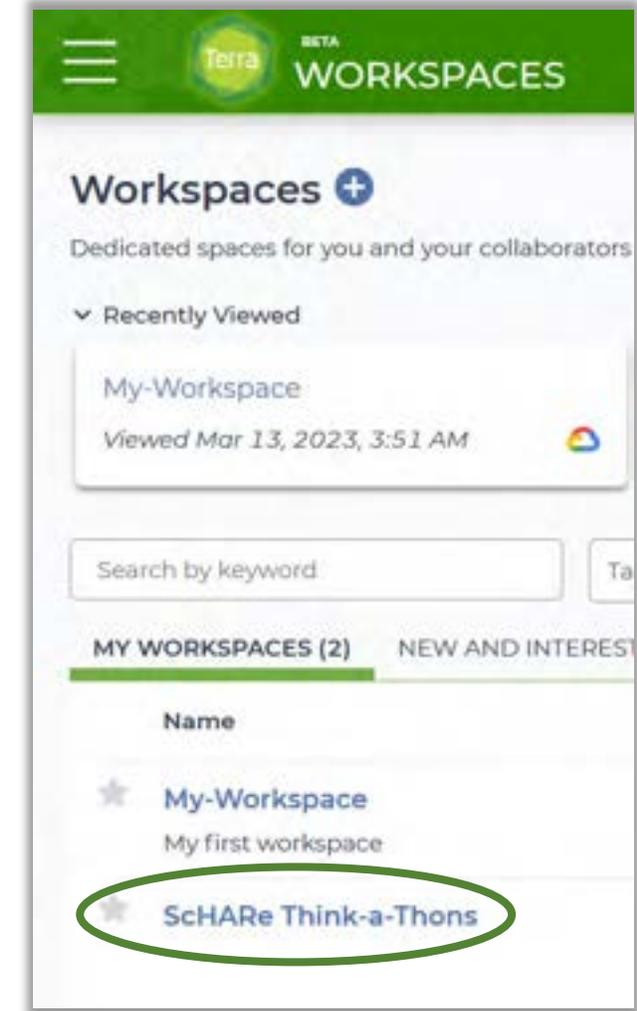
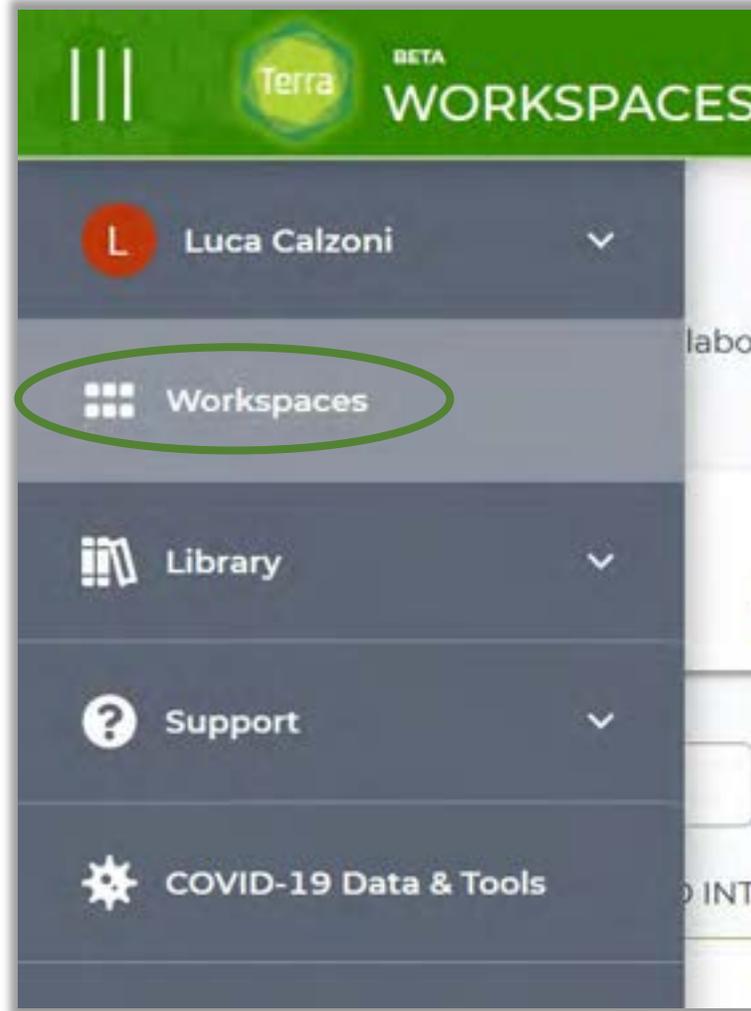
- Creating the cloud environment **will take 3-5 minutes**



Since this newly-created notebook is empty, we will open and run one of the **ScHARe instructional notebooks** instead, to give you a closer look at **how notebooks work**.

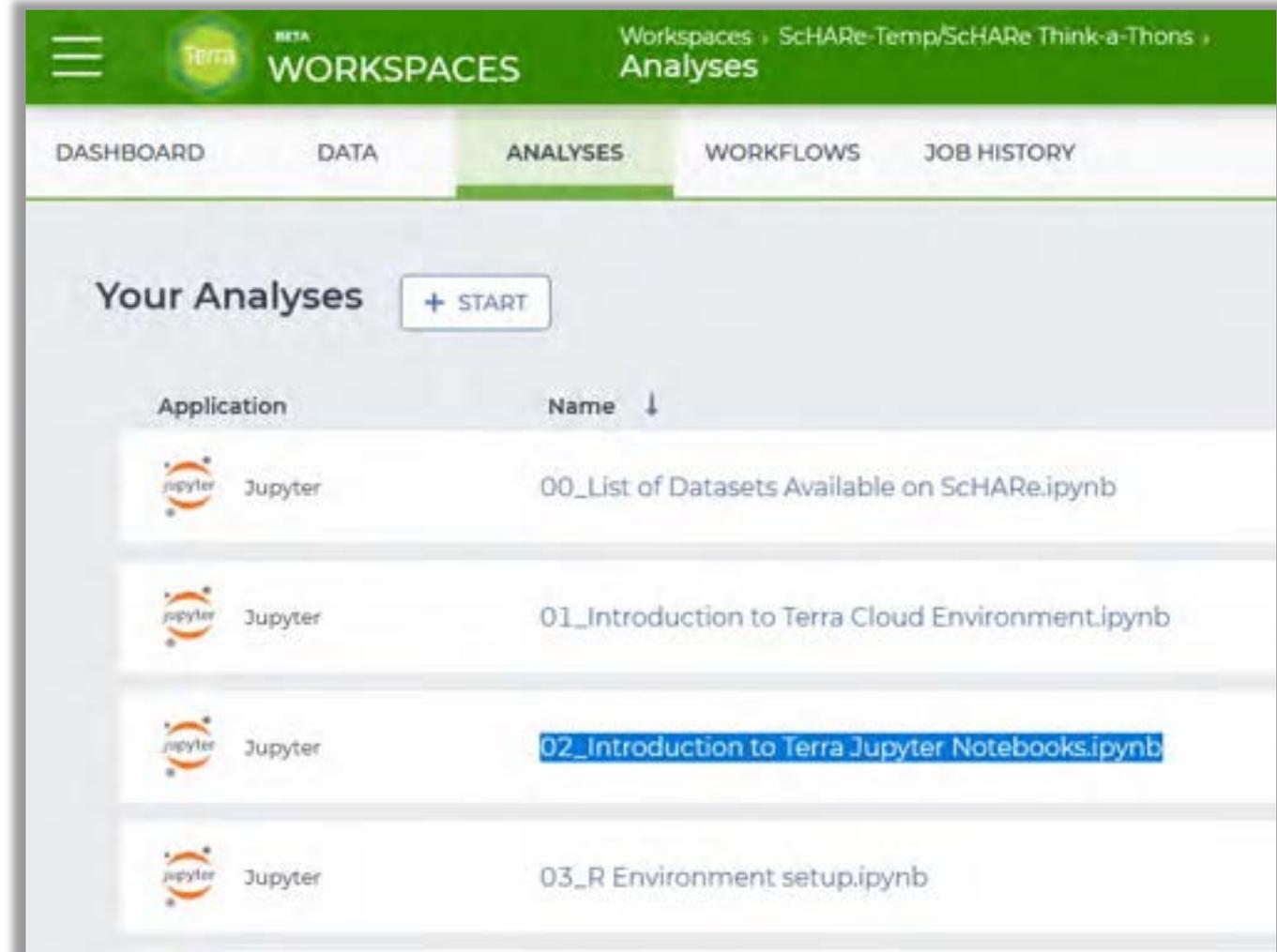
Running a notebook

1. Click on the menu in the top left corner of the page, then on “Workspaces”
2. Click on the ScHARe Think-a-Thons workspace that was shared with you by the instructors



Running a notebook

3. Click on the “02_Introduction to Terra Jupyter Notebooks” notebook



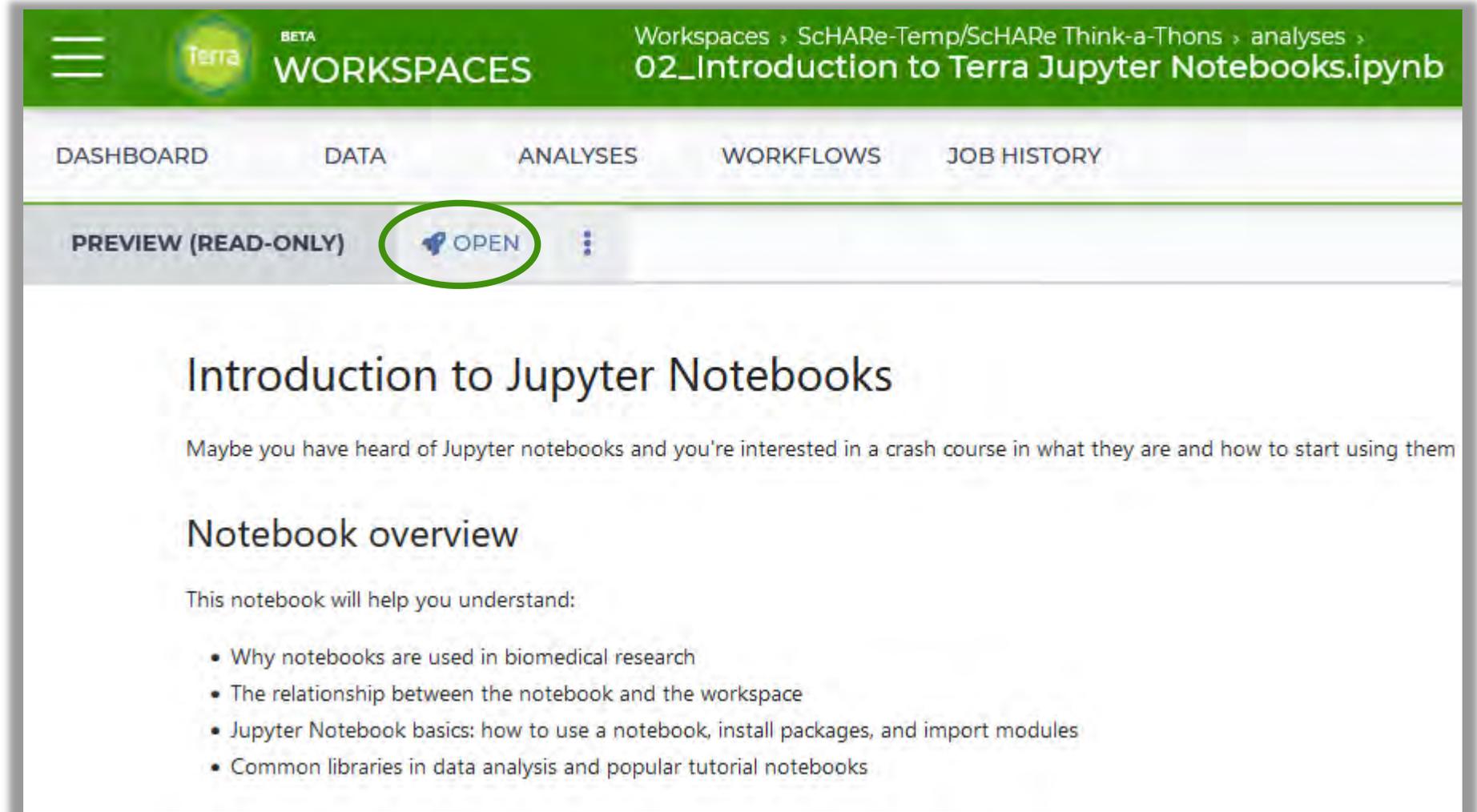
The screenshot displays the Terra WORKSPACES interface. At the top, there is a green header with the Terra logo, the word "WORKSPACES", and the text "BETA". To the right of the header, it shows the current workspace: "Workspaces > ScHARe-Temp/ScHARe Think-a-Thons > Analyses". Below the header is a navigation bar with tabs for "DASHBOARD", "DATA", "ANALYSES" (which is selected), "WORKFLOWS", and "JOB HISTORY".

The main content area is titled "Your Analyses" and includes a "+ START" button. Below this is a table listing the notebooks:

Application	Name ↓
 Jupyter	00_List of Datasets Available on ScHARe.ipynb
 Jupyter	01_Introduction to Terra Cloud Environment.ipynb
 Jupyter	02_Introduction to Terra Jupyter Notebooks.ipynb
 Jupyter	03_R Environment setup.ipynb

Running a notebook

4. Click on
“Open”



The screenshot shows the Terra WORKSPACES interface. The top navigation bar is green and contains the Terra logo, the word "WORKSPACES" in white, and the breadcrumb path: "Workspaces > ScHARe-Temp/ScHARe Think-a-Thons > analyses > 02_Introduction to Terra Jupyter Notebooks.ipynb". Below this is a white navigation bar with tabs for "DASHBOARD", "DATA", "ANALYSES", "WORKFLOWS", and "JOB HISTORY". A grey bar below the navigation tabs shows "PREVIEW (READ-ONLY)" on the left and an "OPEN" button with a blue icon and a vertical ellipsis menu icon on the right. The "OPEN" button is circled in green. The main content area has a white background and contains the following text:

Introduction to Jupyter Notebooks

Maybe you have heard of Jupyter notebooks and you're interested in a crash course in what they are and how to start using them

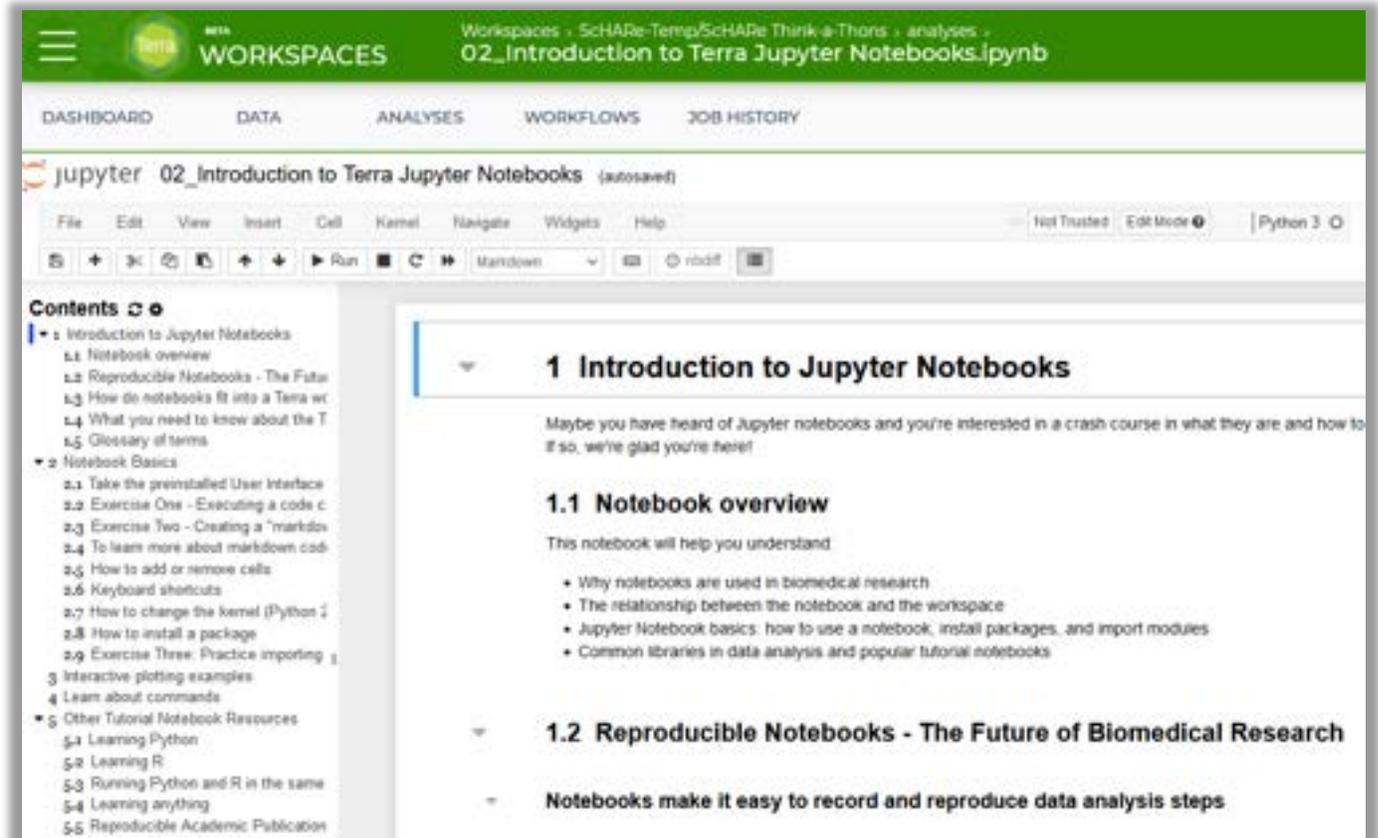
Notebook overview

This notebook will help you understand:

- Why notebooks are used in biomedical research
- The relationship between the notebook and the workspace
- Jupyter Notebook basics: how to use a notebook, install packages, and import modules
- Common libraries in data analysis and popular tutorial notebooks

Running a notebook

5. Configure the environment as shown before, then click on “Create”
6. **Follow the step-by-step instructions in the notebook,** which will help you understand:
 - why notebooks are used in research
 - the relationship between the notebook and the workspace
 - how to use a notebook, install packages, and import modules
 - common libraries in data analysis



The screenshot displays the Jupyter Notebook interface within a workspace. The top navigation bar is green and includes the Terra logo, the word "WORKSPACES", and the current workspace name: "Workspaces - ScHARe-Temp/ScHARe Think-a-Thons - analyses - 02_Introduction to Terra Jupyter Notebooks.ipynb". Below this, a secondary navigation bar contains tabs for "DASHBOARD", "DATA", "ANALYSES", "WORKFLOWS", and "JOB HISTORY". The main interface features a Jupyter menu bar with options like "File", "Edit", "View", "Insert", "Cell", "Kernel", "Navigate", "Widgets", and "Help". A toolbar below the menu bar includes icons for file operations and a "Run" button. On the left, a "Contents" sidebar lists the notebook's sections, such as "Introduction to Jupyter Notebooks" and "Notebook Basics". The main content area shows the first section, "1 Introduction to Jupyter Notebooks", with a sub-section "1.1 Notebook overview" containing a bulleted list of topics.

The logo for SCIARe features the word in a white, bold, sans-serif font. The letters 'C' and 'A' are partially obscured by a stylized orange and yellow cloud. A purple arrow points from the 'C' towards the 'A', and another purple arrow points from the 'A' towards the 'R'.

SCIARe

Part V
Billing and Costs

What are the cloud costs of working on Terra?

The Terra platform is **free to use**

However, the following operations in Terra **may incur charges**:

1. Virtual Machine compute costs

In cloud computing, a **virtual machine** is an emulation of a computer system that provides the functionality of a physical computer

Terra allows you to **customize** the characteristics of your virtual machine based on your computation needs (more on this later)

- A **high-performance machine costs more**
- You will be charged for the **time you use the machine**

The screenshot displays the configuration options for a cloud compute profile. The 'Cloud compute profile' section includes 'CPUs' set to 1 and 'Memory (GB)' set to 3.75. There is an unchecked checkbox for 'Enable GPUs' with a 'BETA' label and a link to learn more about GPU cost and restrictions. The 'Compute type' is set to 'Standard VM'. The 'Enable autopause' checkbox is checked, with a '15' minute inactivity timer and a link to learn more about autopause. The 'Location' is set to 'us-central1 (Iowa) (default)' with a 'BETA' label and an information icon. The 'Persistent disk' section includes a description: 'Persistent disks store analysis data. Learn more about persistent disks and when your disk is mounted.' Below this, the 'Disk Type' is set to 'Standard' and the 'Disk Size (GB)' is set to 10.

What are the cloud costs of working on Terra?

2. Data storage

- You will be charged for any data stored in the storage spaces (“**buckets**”) associated with your account

3. Data egress (i.e. moving data) costs

- As we will see later, when creating a bucket to store data you are asked to set its location (“**region**”), i.e. the physical place where the data is going to reside (e.g.: “US-CENTRAL1 – Iowa”, or “US-EAST4 - Northern Virginia”). More info on regions [here](#). You will pay to **move stored data to a bucket in a different region**
- You will pay to **download data from a bucket that is configured as a Requester Pays bucket**. With such buckets, the data requester pays the cost of the data download instead of the bucket owner

How will I be charged for these costs?

Terra runs on Google Cloud Platform (GCP). All Terra costs are GCP fees that are ultimately paid for by a **Google Cloud Billing account** linked to Terra – specifically, to a **Terra Billing project**

- ▶ Each Billing project is linked to an umbrella Google **Cloud Billing account**
- ▶ A **Terra Billing project** is a pass-through assigned to a workspace when you create it
- ▶ All GCP fees (storage, compute, egress) are charged **per workspace** - *regardless of who does the analysis or whether they have access to a billing project.*



How will I be charged for these costs?

Will I incur any costs today?

Today and for one day after the Think-a-Thon, **access to a free temporary billing project** will allow you to run all the materials with your instructors

What happens after tomorrow?

You will no longer have access to the free temporary billing project. If you want to access work-in-progress from the Think-a-Thon, you will need to **set up your own billing** and copy any of your workspaces to your own billing

Next, we will show you how to set up your own billing

Get \$300 in free Google Cloud credits

If you've never used Google Cloud before, **you are eligible for \$300 in free Google Cloud credits** you can use for working in Terra

Conditions for Google Cloud credits eligibility

- You haven't previously signed up for the Free Trial
- You've never been a paying customer of Google Cloud, Google Maps Platform, or Firebase
- If you're part of an organization that uses Google Cloud, your email will likely not be eligible



What can I do with my credits in Terra?

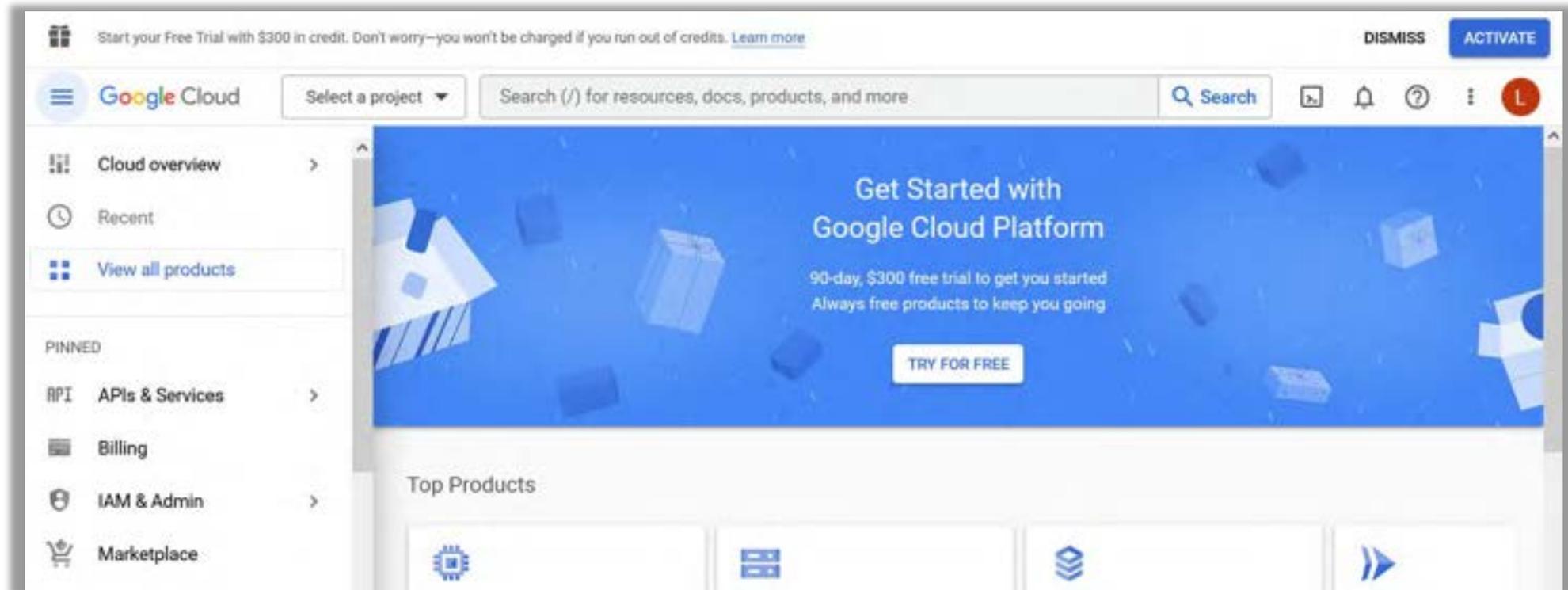
The credits will cover anything that has a cost in Terra - such as storing data and running analyses. You can't use credits to add GPUs to your computing resources, and you are limited to 4 workspaces at a time

How long will my \$300 credits be available?

Your credits will be available for 3 months, or until you have used up all \$300. Once your credits run out or expire, you can upgrade to a paid account

Step 1. Set up a Google Cloud Billing account

1. Go to the **Google Cloud console** at <https://console.cloud.google.com/> and sign in with your Terra user ID. If you haven't already set up a billing account, you'll be invited to activate your free trial
2. Click the **activate** button and follow the instructions



Step 1. Set up a Google Cloud Billing account

Step 1 of 2 Account Information

SWITCH ACCOUNT

Country

United States ▼

What best describes your organization or needs?

Please select
Other ▼

Terms of Service

I have read and agree to the [Google Cloud Platform Terms of Service](#), [Supplemental Free Trial Terms of Service](#), and the terms of service of [any applicable services and APIs](#).

Required to continue

CONTINUE

Access to all Cloud Platform Products

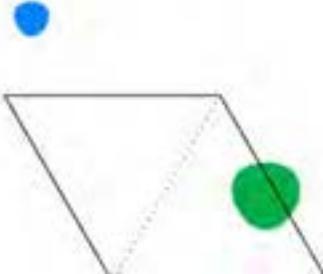
Get everything you need to build and run your apps, websites and services, including Firebase and the Google Maps API.

\$300 credit for free

Put Google Cloud to work with \$300 in credit to spend over the next 90 days.

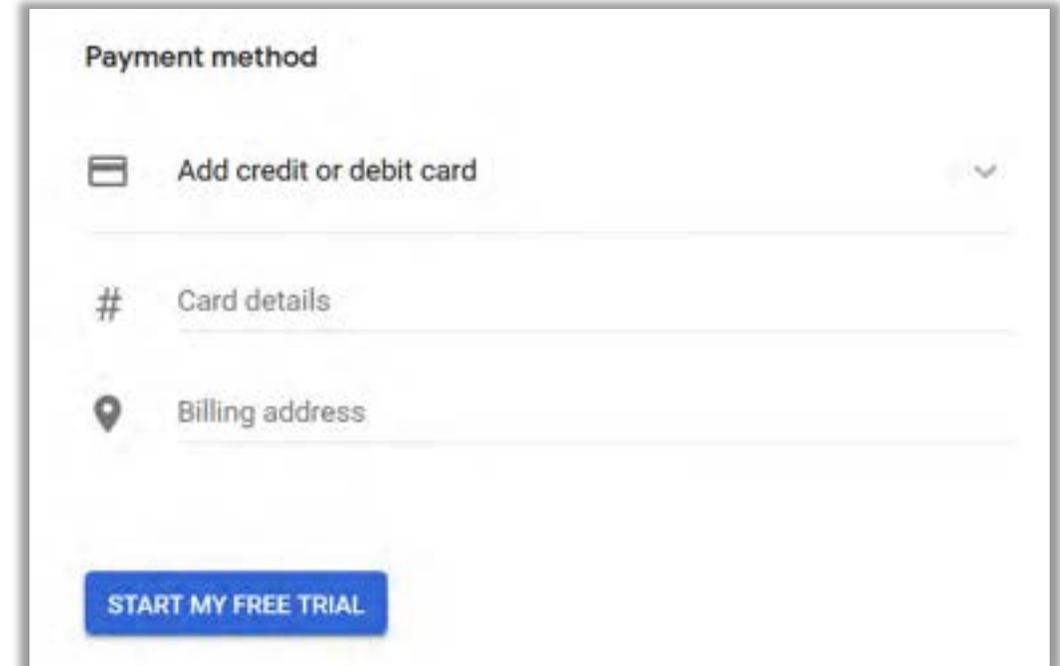
No autocharge after free trial ends

We ask you for your credit card to make sure you are not a robot. You won't be charged unless you manually upgrade to a paid account.



Step 1. Set up a Google Cloud Billing account

3. You'll need to verify your identity with a one-time verification sent to a cell phone, and give a credit card, PayPal account or bank account **(you won't be billed until the credits expire)**
4. **Verify** the Google Cloud Billing account in the [Billing page](#). You should see **My Billing Account** in the top left
5. Google will create a project, **My First Project**, funded by your free credits, in the **My Projects** tab



The screenshot shows the 'Payment method' section of the Google Cloud Billing setup interface. It features three main options: 'Add credit or debit card' (with a card icon and a dropdown arrow), 'Card details' (with a hash icon), and 'Billing address' (with a location pin icon). At the bottom of the section is a prominent blue button labeled 'START MY FREE TRIAL'.

Step 2. Link the Cloud Billing account to Terra

The next step is to link the Google Cloud Billing Account to your Terra account, so that Terra and Google can communicate about cost and billing

You must use the same Google ID for both the Cloud Billing account and your Terra user name

1. When logged into Google with your Terra user ID, go to the [Google Cloud Console Billing page](#).
2. Select the **checkbox beside the Google Cloud billing account** you will use for Terra.
3. On the right panel, below **Permissions**, select the **Add Principal** button.
4. Add "**terra-billing@terra.bio**" under **New Principal** in the form.
5. In the dropdown, select the role **Billing > Billing Account User**.
6. Click **Add**.
7. Click on the **Save** button

Note: "terra-billing@terra.bio" will appear in the list as "terra-billing@firecloud.org." This is expected.

Step 3. Create a Terra Billing project

Once Terra is linked to a Cloud Billing account, you can create a Terra Billing project that will allow you to create a workspace to store and analyze data

1. Go to the [Billing page](#) from the main navigation (click on **your name** to expand the drop-down, and select **Billing**)
2. Click on the **"+ Create"** button at the top left
3. If prompted to **Enable Billing Permissions**, select the **Google identity** of the Google Cloud Billing account, and click **Allow**. This lets Terra access Cloud Billing accounts associated with your Terra user name (Google ID).
4. Enter a **unique name** for your Terra Billing project
5. Select the **Google Cloud Billing account** that will fund the Billing project

You may see multiple Cloud Billing accounts that you can select for this Terra Billing project. If you need to locate a Billing account ID, navigate to the [Google Developers Console](#) and click on **Billing**. Look for the number below **Billing account ID**

Understanding and monitoring costs

You can **ESTIMATE COSTS**:

1. **analysis costs**
2. cloud storage costs
3. egress (i.e., data moving) costs

You can **CHECK ACTUAL COSTS** in the Google Cloud Platform Console

You can **REDUCE COSTS** in several ways (for advanced users)

Cloud Environment

A cloud environment consists of application configuration, cloud compute and persistent disk(s).

Running cloud compute cost	Recent cloud compute cost	Persistent disk cost
\$0.06 per hr	< \$0.01 per hr	\$2.00 per month

Application configuration

Default: (DATA 4.2.0.0, Python 3.7.10, W 4.1.1)

What's installed on this environment? Updated: Sep 23, 2021 Version: 2.0.1

Cloud compute profile

CPUs: 1 Memory (GB): 175

Enable GPUs [Learn more about GPU cost and restrictions.](#)

Startup script

URI

Compute type

Standard VM

Persistent disk size (GB)

Persistent disks store analytics data. [Learn more about persistent disks and where your disk is mounted.](#)

30

Cloud Environment

RUNNING (\$0.20 hr)

- 1) *Adjust settings to optimize cost (VM and disk)*
- 2) *Estimate costs using real-time cost/hour in Cloud Environment widget*

- ▶ Updates based on the machine configuration you choose
- ▶ Total cost (estimate) = (cost/hour) x (hours the VM will be active) + cost of the Persistent Disk
- ▶ Autopause function saves money!

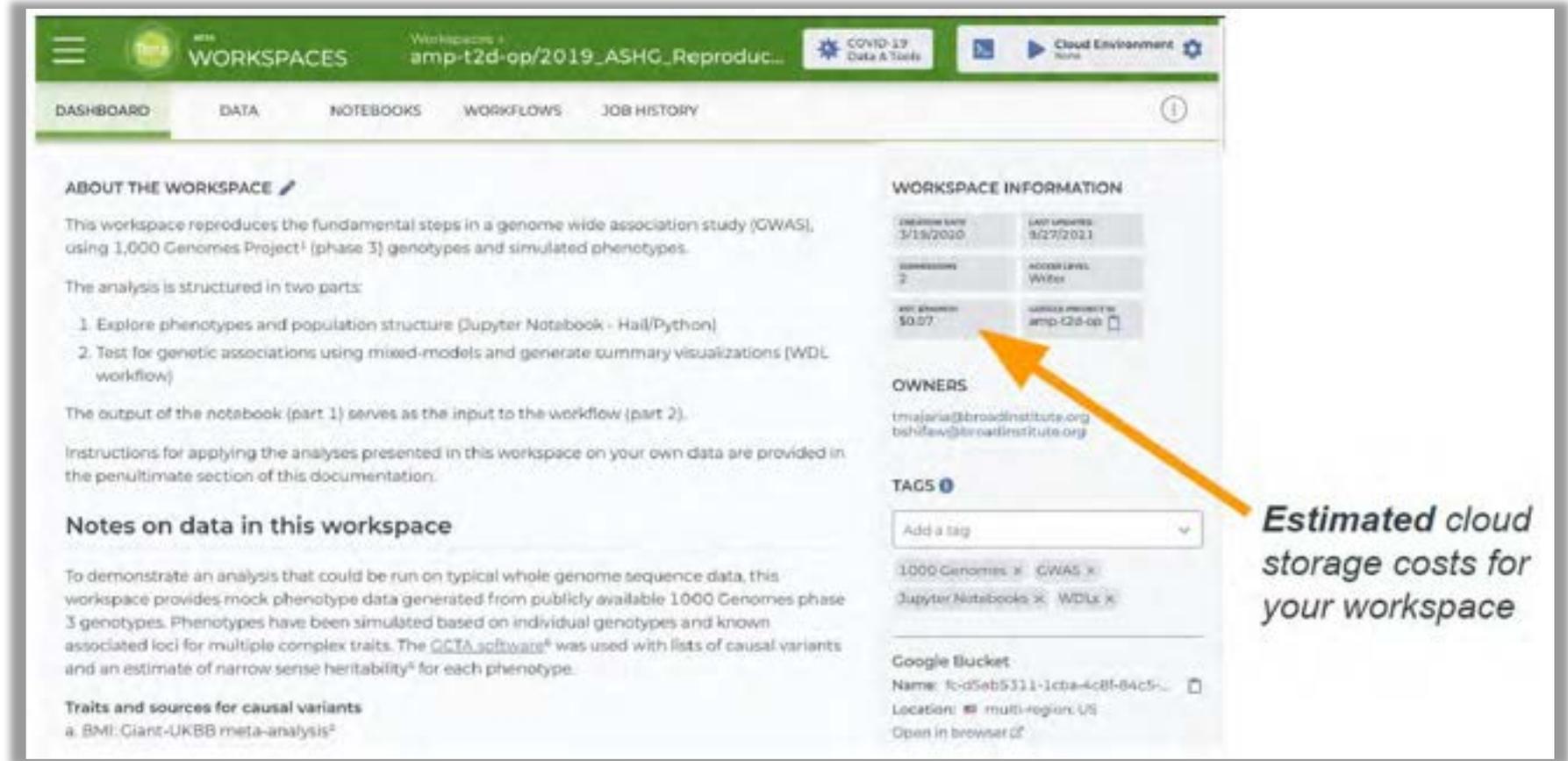
Understanding and monitoring costs

You can **ESTIMATE COSTS**:

1. analysis costs
2. **cloud storage costs**
3. egress (i.e., data moving) costs

You can **CHECK ACTUAL COSTS** in the Google Cloud Platform Console

You can **REDUCE COSTS** in several ways (for advanced users)



The screenshot displays the Google Cloud Platform Workspace interface for a workspace named "amp-t2d-op/2019_ASHG_Reproduc...". The interface includes a navigation bar with "DASHBOARD", "DATA", "NOTEBOOKS", "WORKFLOWS", and "JOB HISTORY". The main content area is divided into two columns. The left column, titled "ABOUT THE WORKSPACE", provides details about the workspace's purpose (reproducing GWAS steps) and its structure. The right column, titled "WORKSPACE INFORMATION", contains a table with the following data:

CREATION DATE	LAST UPDATED
3/13/2020	9/27/2021
MEMBERS	ACCESS LEVEL
2	Writer
EST. STORAGE	STORAGE PROVISIONED
\$0.07	amp-t2d-op

An orange arrow points from the text "Estimated cloud storage costs for your workspace" to the "\$0.07" value in the "EST. STORAGE" row. Below the table, the "OWNERS" section lists "t.majara@broadinstitute.org" and "b.shifaw@broadinstitute.org". The "TAGS" section includes a search box and tags for "1000 Genomes", "GWAS", "Jupyter Notebooks", and "WDLs". At the bottom, the "Google Bucket" section shows the name "fc-d5eb5311-1cbe-4c8f-84c5-..." and location "multi-region: US".

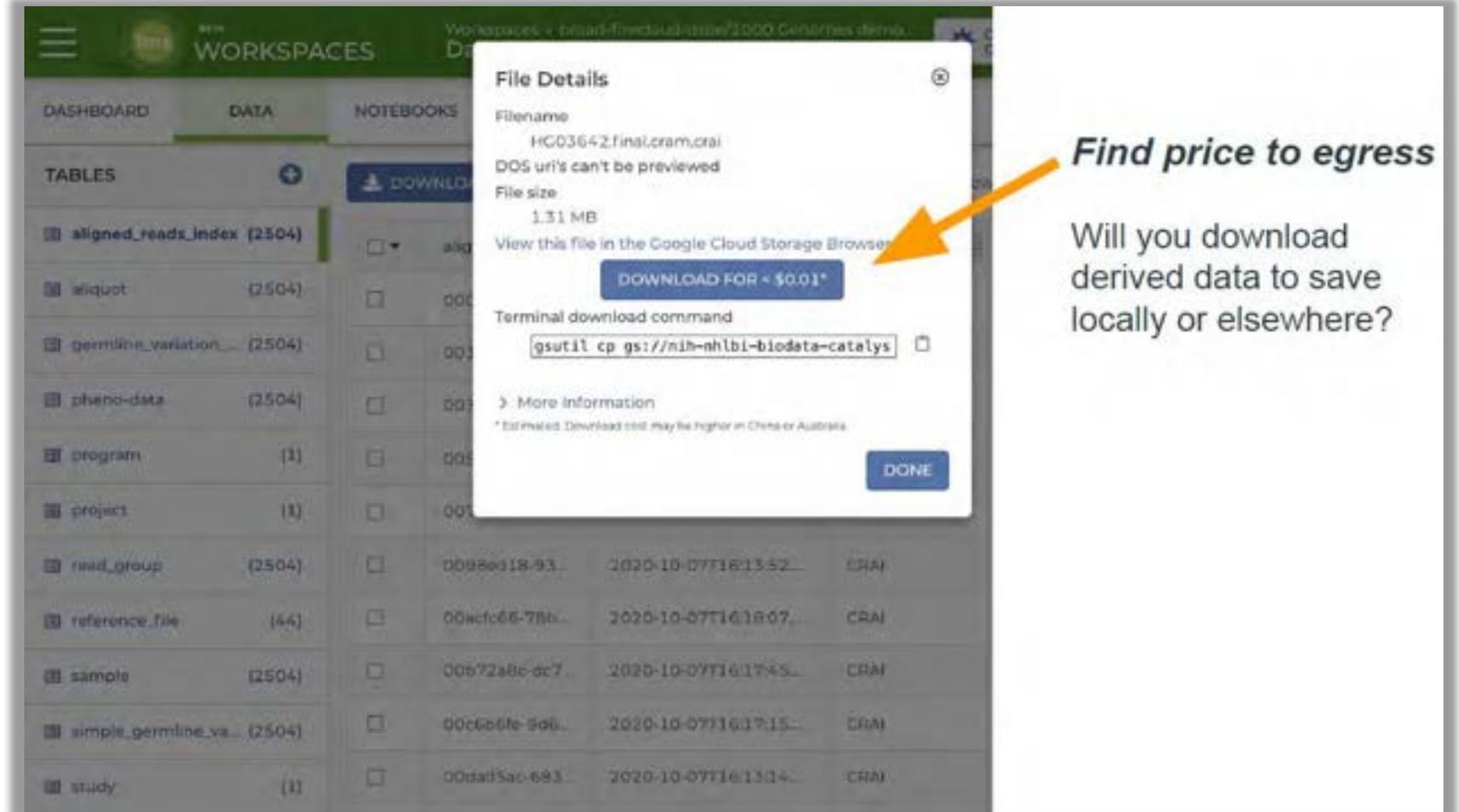
Understanding and monitoring costs

You can **ESTIMATE COSTS**:

1. analysis costs
2. cloud storage costs
3. egress (i.e., data moving) costs

You can **CHECK ACTUAL COSTS** in the Google Cloud Platform Console

You can **REDUCE COSTS** in several ways (for advanced users)



File Details

Filename
HG03642.final.cram.crai

DOS uri's can't be previewed

File size
1.31 MB

View this file in the Google Cloud Storage Browser

DOWNLOAD FOR < \$0.01*

Terminal download command
`gsutil cp gs://nih-nhlbi-biodata-catalys`

> More Information

* Estimated. Download cost may be higher in China or Australia.

DONE

Find price to egress

Will you download derived data to save locally or elsewhere?

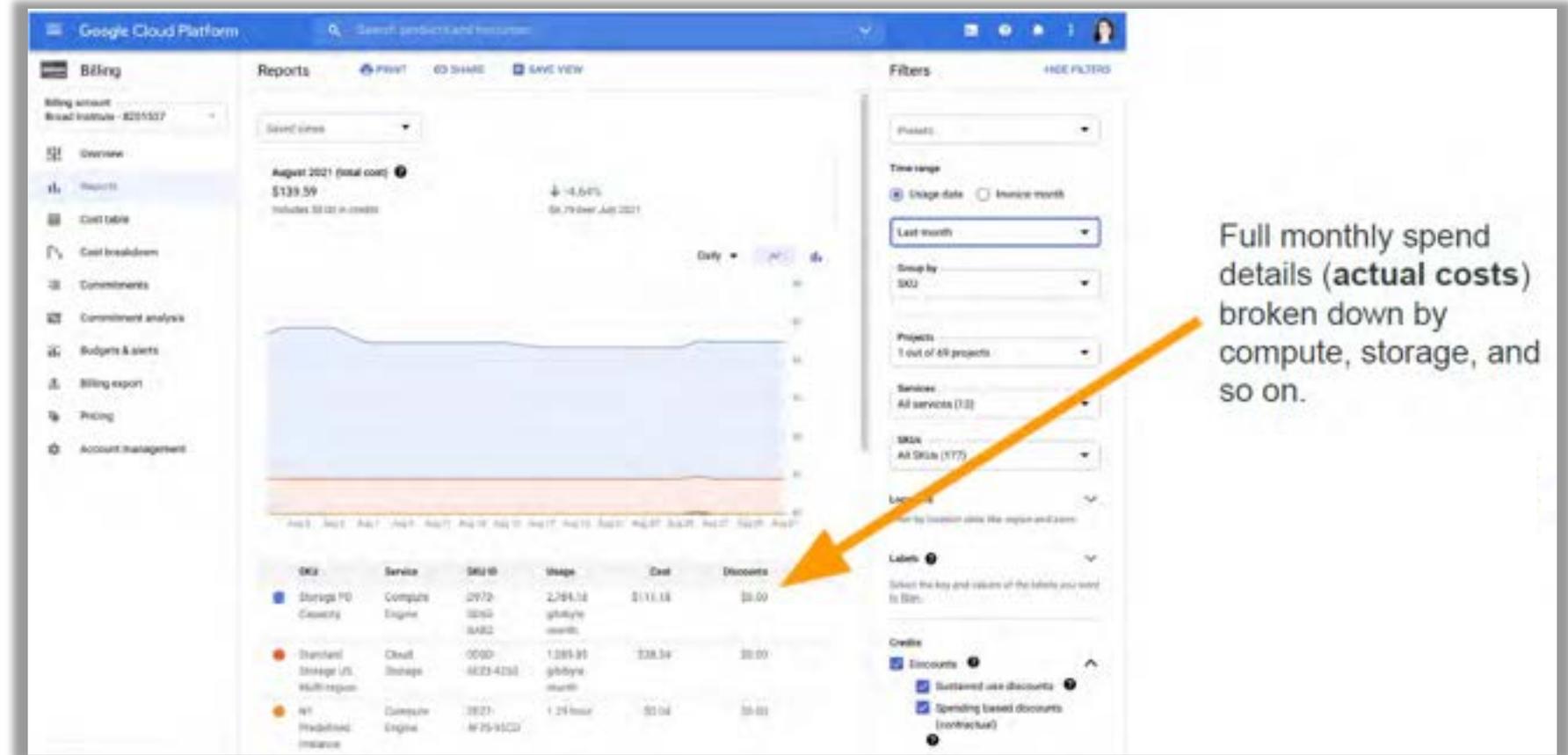
Understanding and monitoring costs

You can **ESTIMATE COSTS**:

1. analysis costs
2. cloud storage costs
3. egress (i.e., data moving) costs

You can **CHECK ACTUAL COSTS** in the Google Cloud Platform Console

You can **REDUCE COSTS** in several ways (for advanced users)



Understanding and monitoring costs

You can **ESTIMATE COSTS**:

1. analysis costs
2. cloud storage costs
3. egress (i.e., data moving) costs

You can **CHECK ACTUAL COSTS** in the Google Cloud Platform Console

You can **REDUCE COSTS** in several ways (guides are for advanced users)

Terra allows you to find the right balance between cost and time

Saving on workflow costs

- ▶ Delete intermediate files: [guide](#)
- ▶ Call-caching: [guide](#)
- ▶ Checkpointing: [guide](#)
- ▶ Preemptible VMs: [guide](#)

Saving Cloud Environment costs

- ▶ Size application compute appropriately: [guide](#)
- ▶ Move generated data to regional or nearline storage: [guide](#)
- ▶ Autopause: [guide](#)

Saving on storage costs

- ▶ Ask how much are you storing, where are you storing it, and how frequently will you access it?
- ▶ Move data to regional or nearline storage: [guide](#)

The word "SCIENCE" is written in a large, white, sans-serif font. The letters "I" and "A" are partially obscured by a stylized orange and yellow cloud. A purple arrow points from the "I" to the "A".

SCIENCE

Thank you

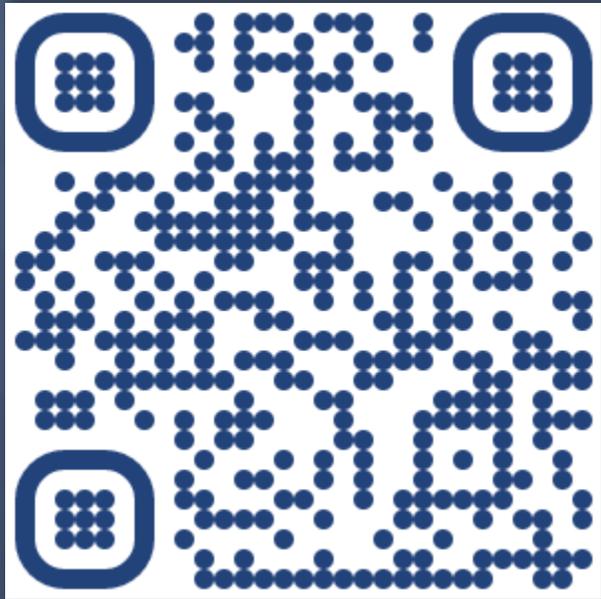
**Next Think-a-Thons:
Data Sets Review and Analyzing Data**

Terra tutorials and resources

If you are new to Terra, we recommend exploring the following resources:

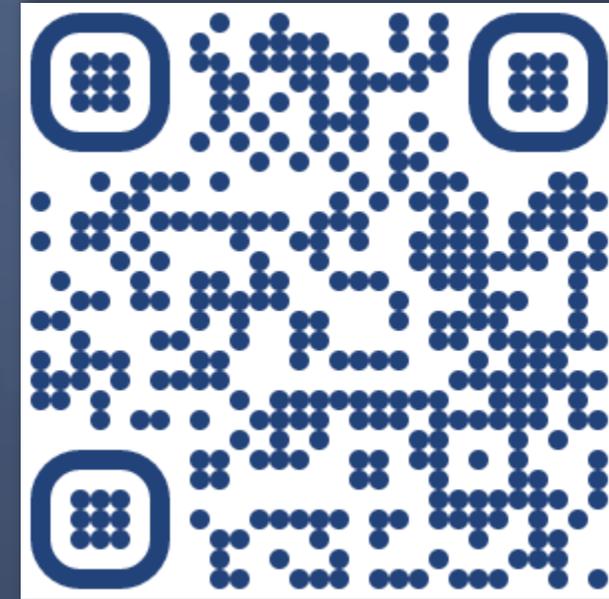
- [Overview Articles](#): Review high-level docs that outline what you can do in Terra, how to set up an account and account billing, and how to access, manage, and analyze data in the cloud
- [Video Guides](#): Watch live demos of the Terra platform's useful features
- [Terra Courses](#): Learn about Terra with free modules on the Leanpub online learning platform
- [Data Tables QuickStart Tutorial](#): Learn what data tables are and how to create, modify, and use them in analyses
- [Notebooks QuickStart Tutorial](#): Learn how to access and visualize data using a notebook
- [Machine Learning Advanced Tutorial](#): Learn how Terra can support machine learning-based analysis

Next Think-a-Thons:



bit.ly/think-a-thons

Register for ScHARe:



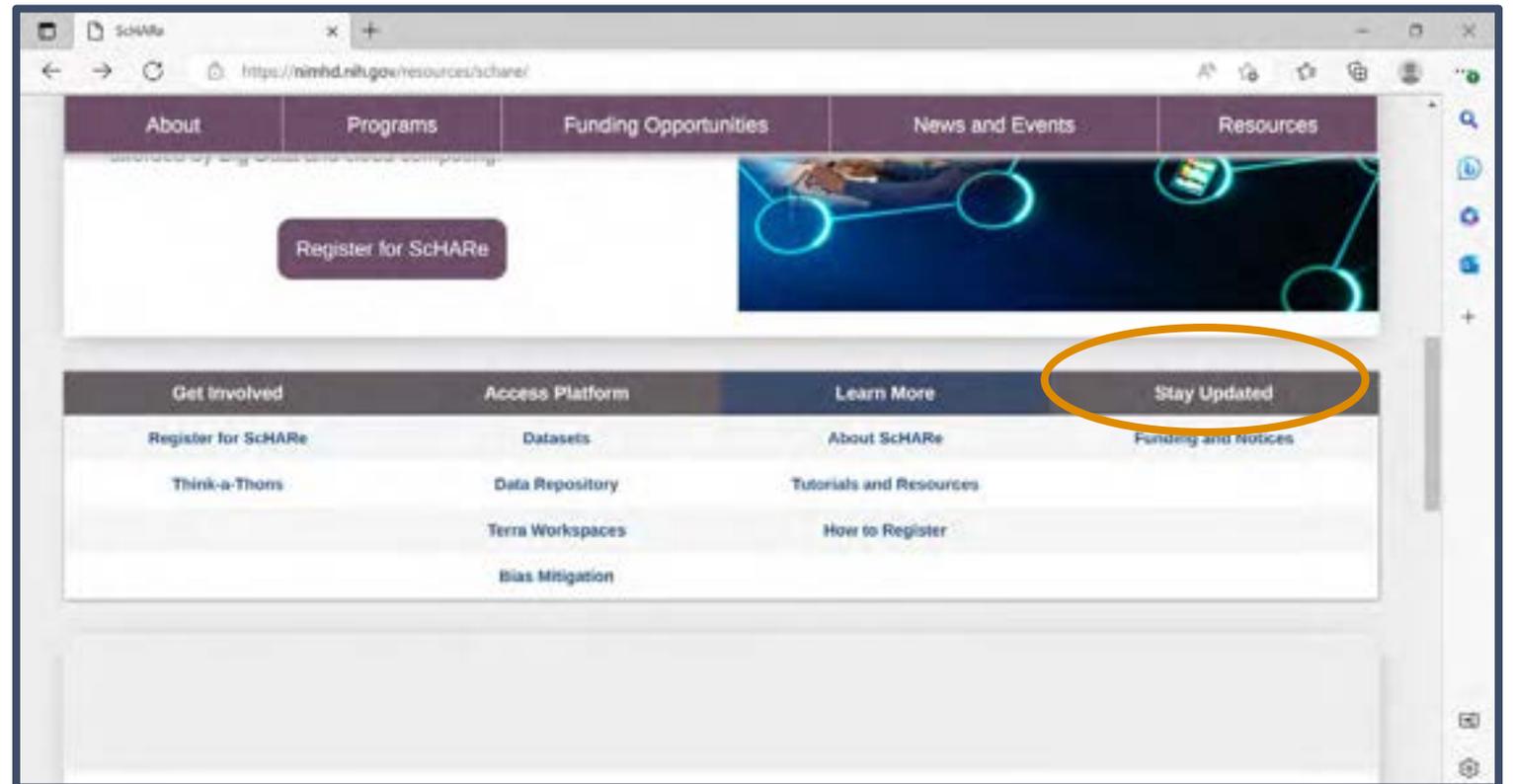
bit.ly/join-schare

✉ schare@mail.nih.gov

Two Ways to Sign up for ScHARe News

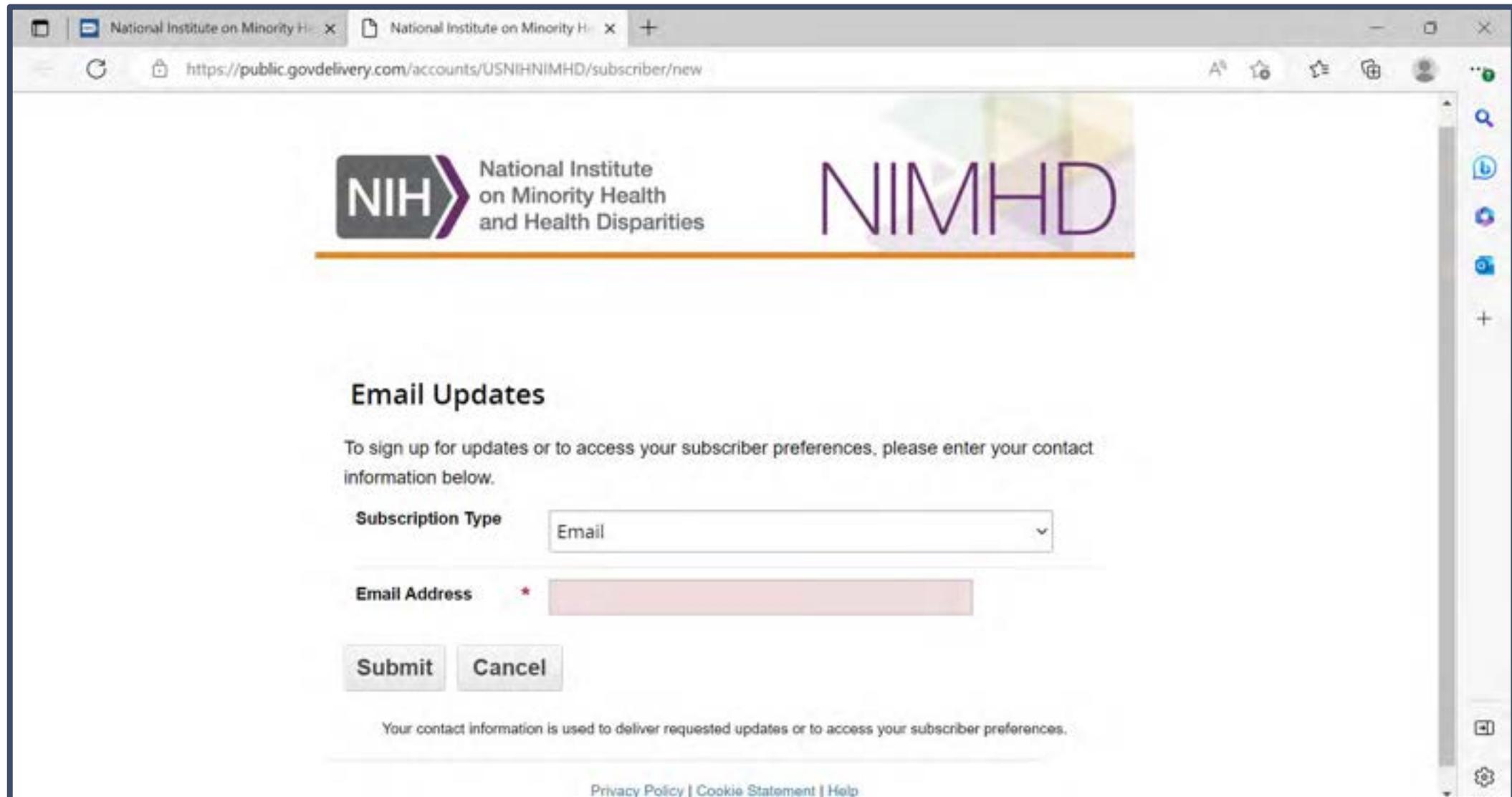


Scannable from your
screen!



nimhd.nih.gov/schare

Step 1: Email Address



The screenshot shows a web browser window with two tabs open, both titled "National Institute on Minority Health and Health Disparities". The address bar displays the URL: <https://public.govdelivery.com/accounts/USNIHNIMHD/subscriber/new>. The page header features the NIH logo on the left and the NIMHD logo on the right. Below the logos, the text "National Institute on Minority Health and Health Disparities" is visible. The main content area is titled "Email Updates" and contains the following text: "To sign up for updates or to access your subscriber preferences, please enter your contact information below." The form includes a "Subscription Type" dropdown menu set to "Email" and an "Email Address" input field with a red asterisk indicating it is required. Below the input field are "Submit" and "Cancel" buttons. At the bottom of the form, a disclaimer states: "Your contact information is used to deliver requested updates or to access your subscriber preferences." At the very bottom of the page, there are links for "Privacy Policy", "Cookie Statement", and "Help".

NIH National Institute on Minority Health and Health Disparities

NIMHD

Email Updates

To sign up for updates or to access your subscriber preferences, please enter your contact information below.

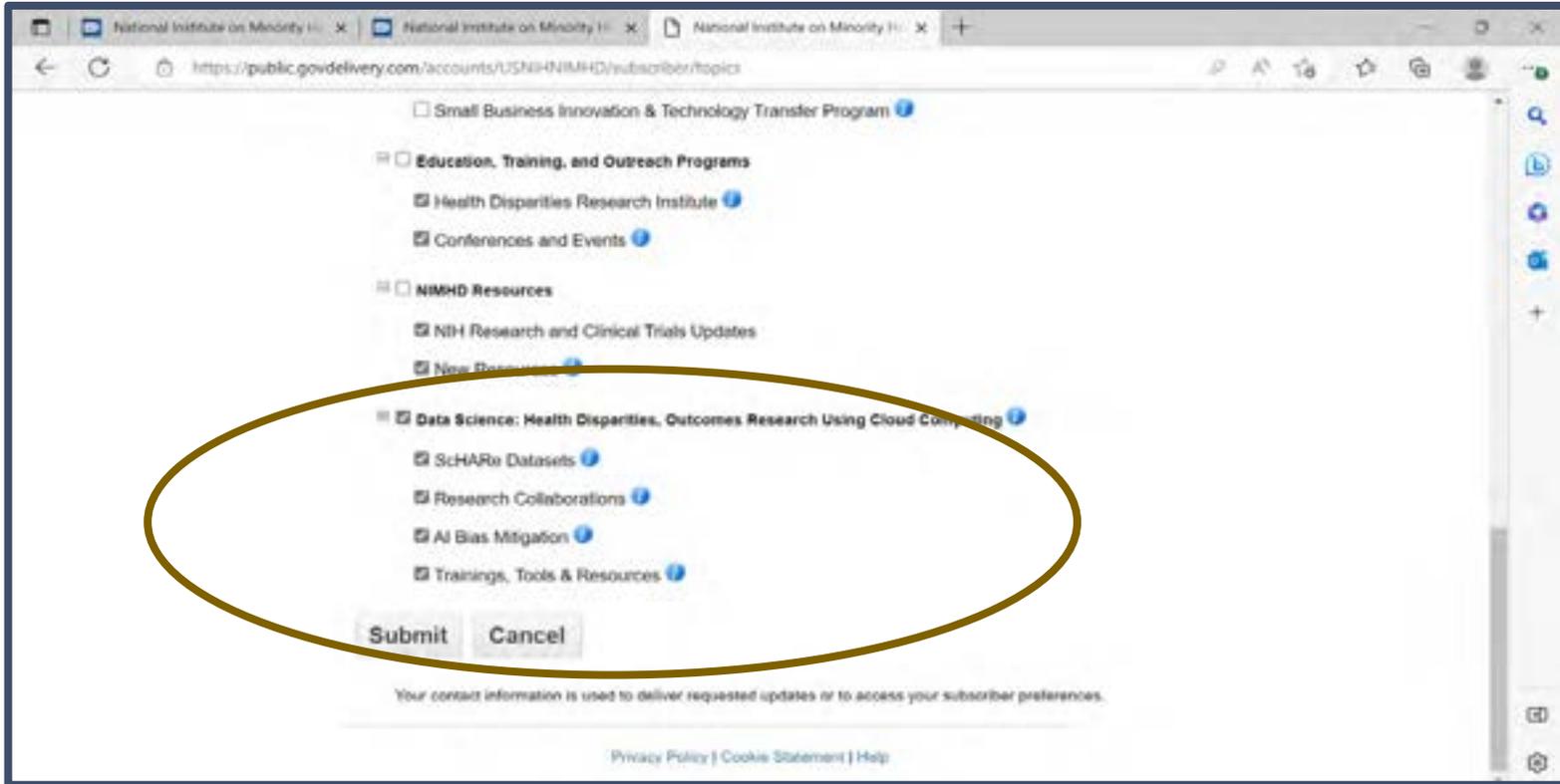
Subscription Type

Email Address *

Your contact information is used to deliver requested updates or to access your subscriber preferences.

[Privacy Policy](#) | [Cookie Statement](#) | [Help](#)

Step 2: Listserv Selection



The screenshot shows a web browser window with the URL <https://public.govdelivery.com/accounts/USNIMHD/subscriber/topics>. The page displays a list of topics for selection, each with a checkbox and a help icon. A yellow oval highlights the following section:

- Data Science: Health Disparities, Outcomes Research Using Cloud Computing**
 - SchARe Datasets
 - Research Collaborations
 - AI Bias Mitigation
 - Trainings, Tools & Resources

At the bottom of the highlighted section are 'Submit' and 'Cancel' buttons. Below the buttons, a small text line reads: "Your contact information is used to deliver requested updates or to access your subscriber preferences." At the very bottom of the page, there are links for "Privacy Policy | Cookie Statement | Help".



Scroll down for SchARe listserv options to sign up for news about:

- **SchARe Datasets** – new datasets coming regularly
- **Research Collaborations** – opportunities to join multidisciplinary SchARe-based research teams
- **AI Bias Mitigation** – bias mitigation tools and research updates
- **Trainings, Tools, & Resources** – Think-a-Thons, teaching resources, & more!

References and credits

- **Tutorials and notebooks:** The Broad Institute, Inc., Verily Life Sciences, LLC