

Durban Exploratory Data Analysis Workshop

22-26 August, 2022

Nelson Mandela School of Medicine

University of KwaZulu-Natal

Overview

- Introductions
- Schedule & Concepts
- Expectations
- Using terra.bio



Workshop Team

Organizers

- Scott Handley – Washington University School of Medicine
- Joseph Elsherbini – Ragon Institute, Harvard University
- Doug Kwon - Ragon Institute, Harvard University
- Blythe Gulley - Ragon Institute, Harvard University

Instructors

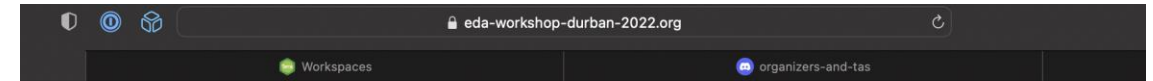
- Lindsay Droit – Washington University School of Medicine
- Fatima Hussain - Ragon Institute, Harvard University

Teaching Assistants

- Derek Tshiabuila – KRISP / CERI
- Marothi Letsoalo – CAPRISA
- Johan van der Molen - CAPRISA

Workshop Command Center

- <https://eda-workshop-durban-2022.org>
 - Logistics
 - Modules
- Schedule



Durban Exploratory Data Analysis Workshop 2022

Home Logistics Modules

Participants please check the Logistics page

Please check out the Logistics and make sure you've completed all steps ahead of our first meeting Monday August 22.

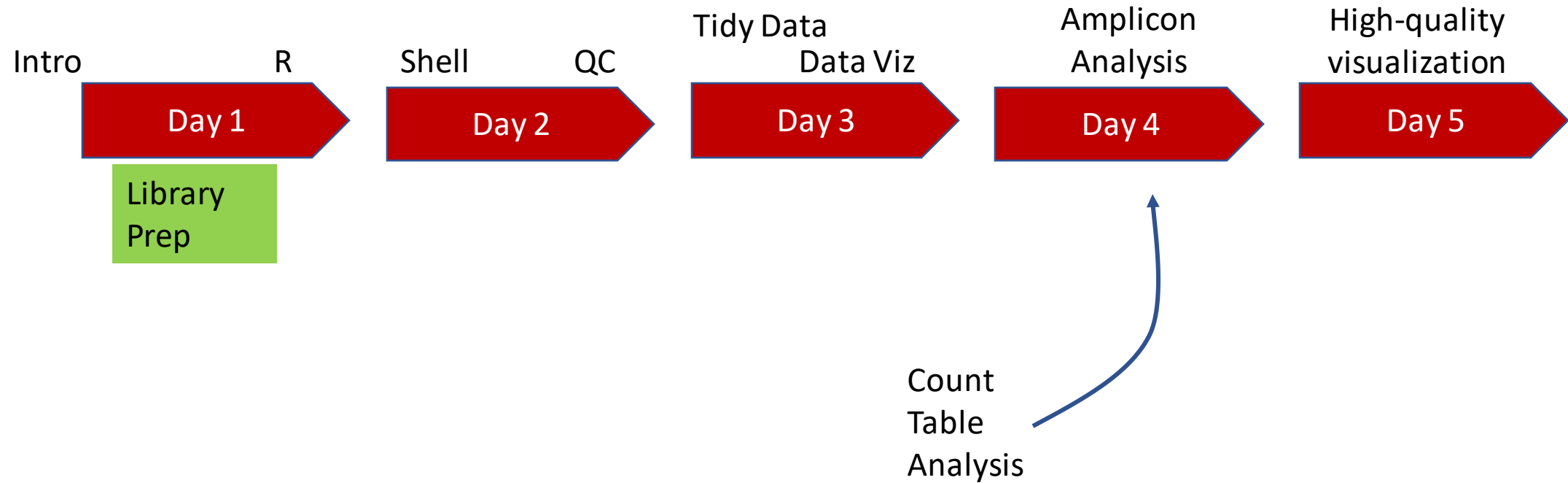


2022 Durban Workshop on Exploratory Data Analysis and Modelling in R

A workshop exploring data analysis and modelling techniques in R with a special emphasis on microbiome applications. This workshop will run August 22-26, 2022 at the Nelson Mandela School of Medicine, University of KwaZulu Natal, in Durban, SA.

Schedule

Schedule & Concepts



Expectations

- There will be some technical issues
 - Be patient please!
- You may not finish every exercise
 - All of the material will be available on-line
- Ask questions!
 - Even if it is not on the schedule
- Help your neighbor
 - There are more of you than there are of us
- The Workshop is designed to establish a foundation for future learning



Focus on your science

Access data, run analysis tools,
and collaborate in Terra:
a scalable platform for biomedical
research.

[GET STARTED](#)

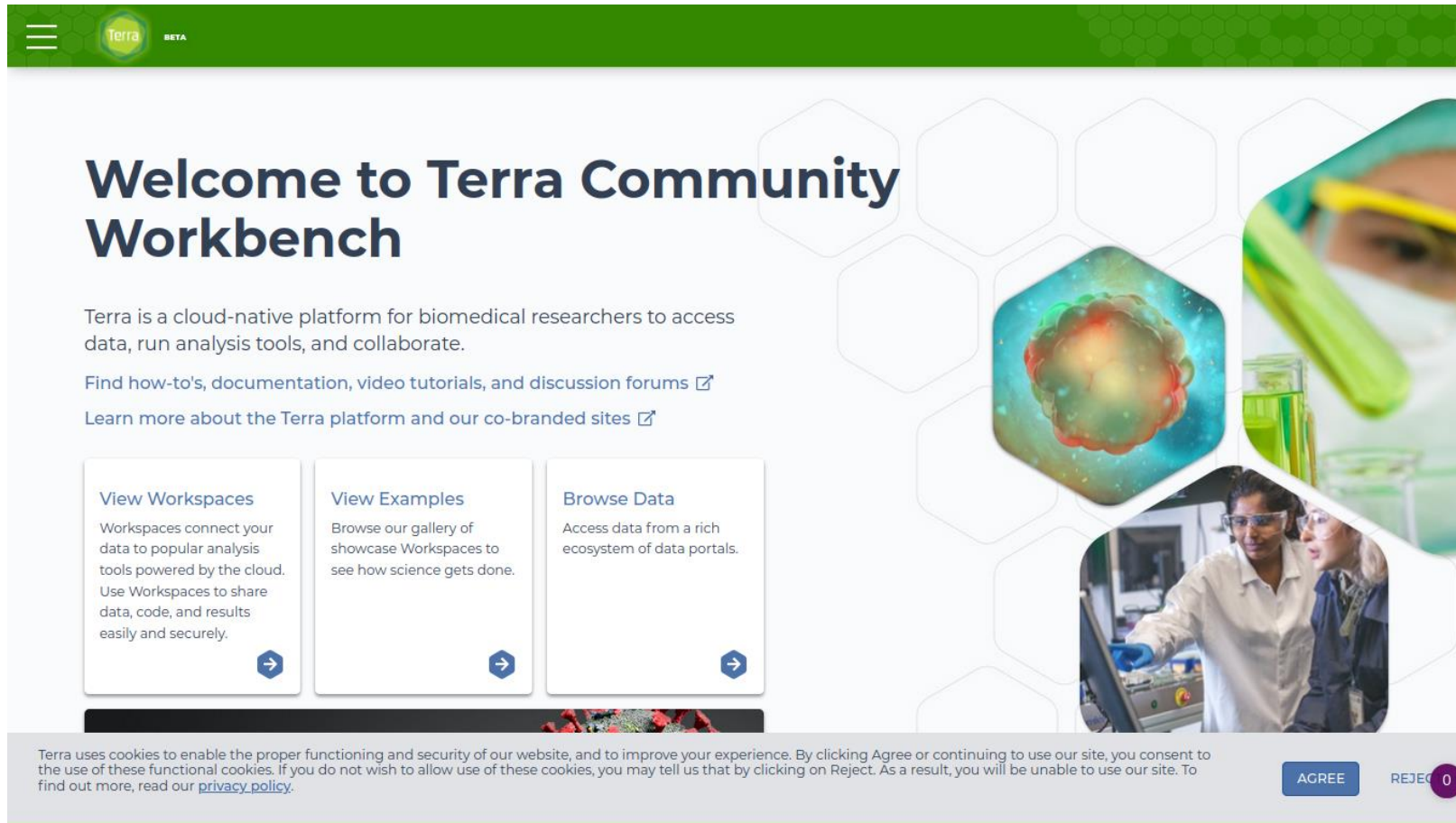
Terra access instructions

25,419
TERRA USERS

2.38 MILLION
STUDY PARTICIPANTS

39.54 MILLION
SINGLE CELLS

Go to the app.terra.bio website



The screenshot shows the homepage of the Terra Community Workbench. The header is green with the Terra logo and a 'BETA' tag. The main heading is 'Welcome to Terra Community Workbench'. Below it, a paragraph describes Terra as a cloud-native platform for biomedical researchers. Two links are provided: 'Find how-to's, documentation, video tutorials, and discussion forums' and 'Learn more about the Terra platform and our co-branded sites'. Three cards are displayed: 'View Workspaces', 'View Examples', and 'Browse Data', each with a brief description and a right-pointing arrow icon. The right side of the page features a collage of images: a hexagonal frame containing a colorful molecular model, a close-up of a person in a lab coat and safety glasses holding a test tube, and a photo of two researchers in a lab setting. At the bottom, a cookie consent banner is visible with 'AGREE' and 'REJECT' buttons.

Welcome to Terra Community Workbench

Terra is a cloud-native platform for biomedical researchers to access data, run analysis tools, and collaborate.

[Find how-to's, documentation, video tutorials, and discussion forums](#)

[Learn more about the Terra platform and our co-branded sites](#)

View Workspaces

Workspaces connect your data to popular analysis tools powered by the cloud. Use Workspaces to share data, code, and results easily and securely.

[→](#)

View Examples

Browse our gallery of showcase Workspaces to see how science gets done.

[→](#)

Browse Data

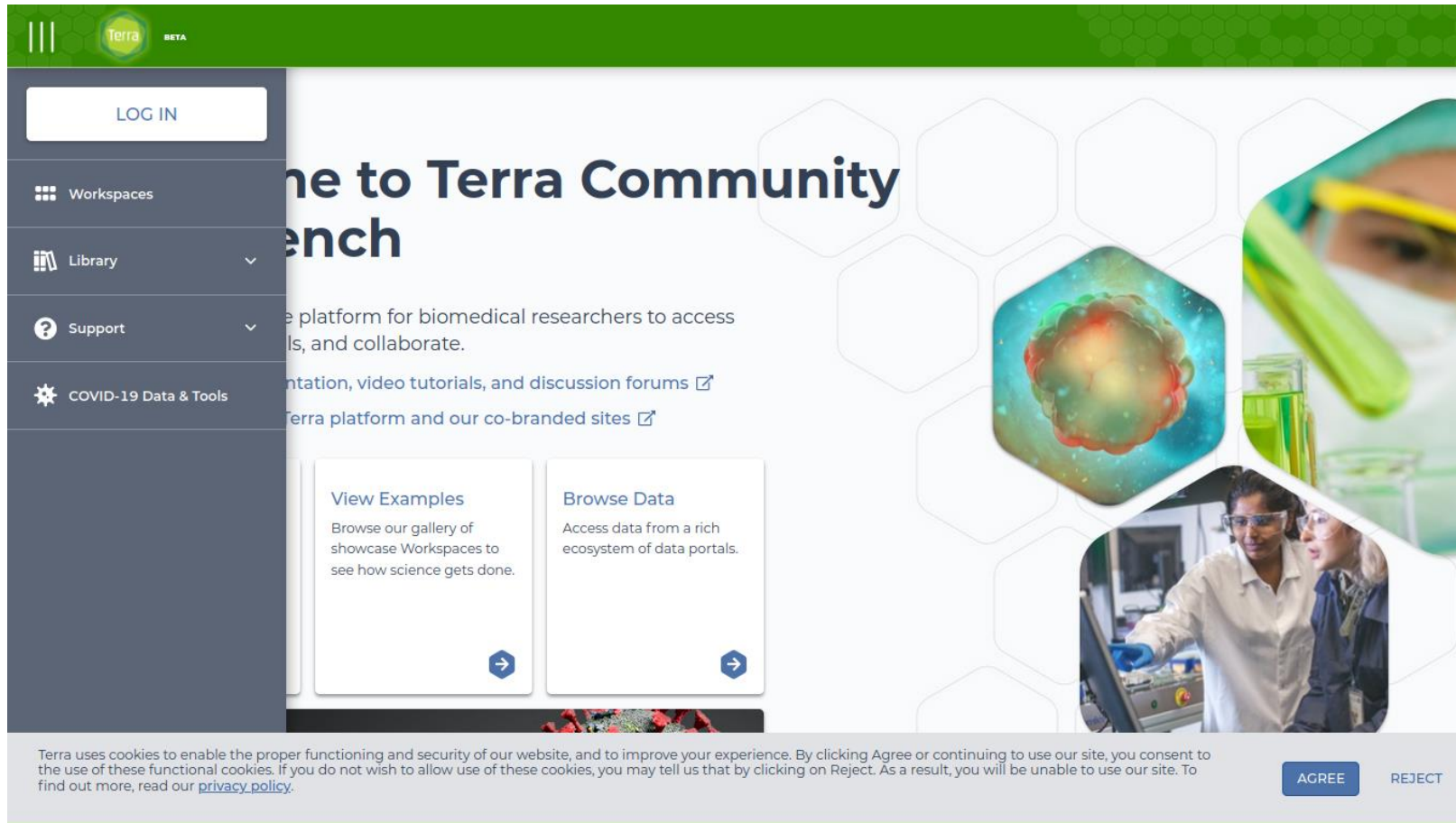
Access data from a rich ecosystem of data portals.

[→](#)

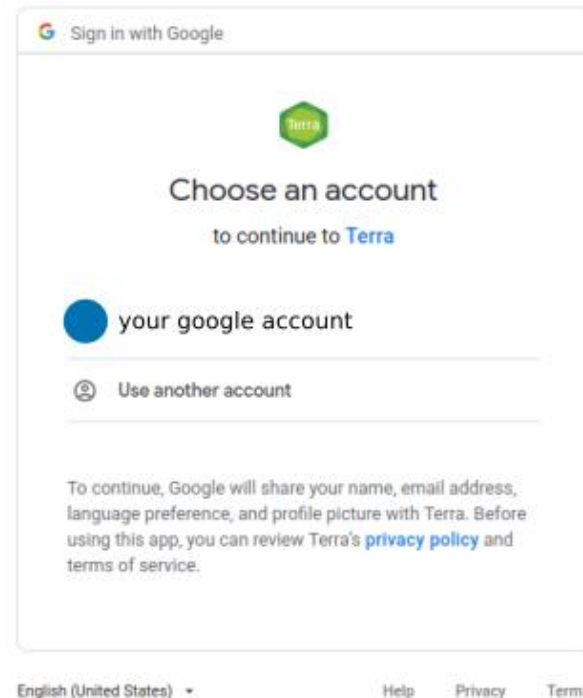
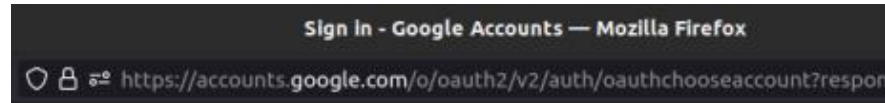
Terra uses cookies to enable the proper functioning and security of our website, and to improve your experience. By clicking Agree or continuing to use our site, you consent to the use of these functional cookies. If you do not wish to allow use of these cookies, you may tell us that by clicking on Reject. As a result, you will be unable to use our site. To find out more, read our [privacy policy](#).

[AGREE](#) [REJECT](#) 0

Click the top left menu and log in



Choose your google account to log in




What if I don't already have a Gmail address or other Google account?

You can follow these steps to create a Google Apps account associated with your existing email address:

1. Go to the Google [sign-up page](#).
2. Under "Choose your username", click "I prefer to use my current email address"
3. With your new Google account set up, you'll be able to [register for Terra with that email address](#).

Make your Terra account



TERRA

New User Registration

First Name *

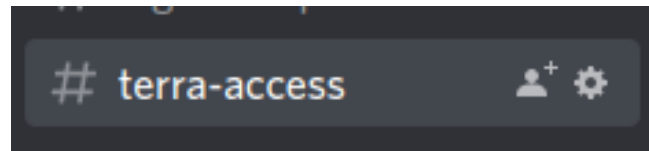
Last Name *

Contact Email for Notifications *

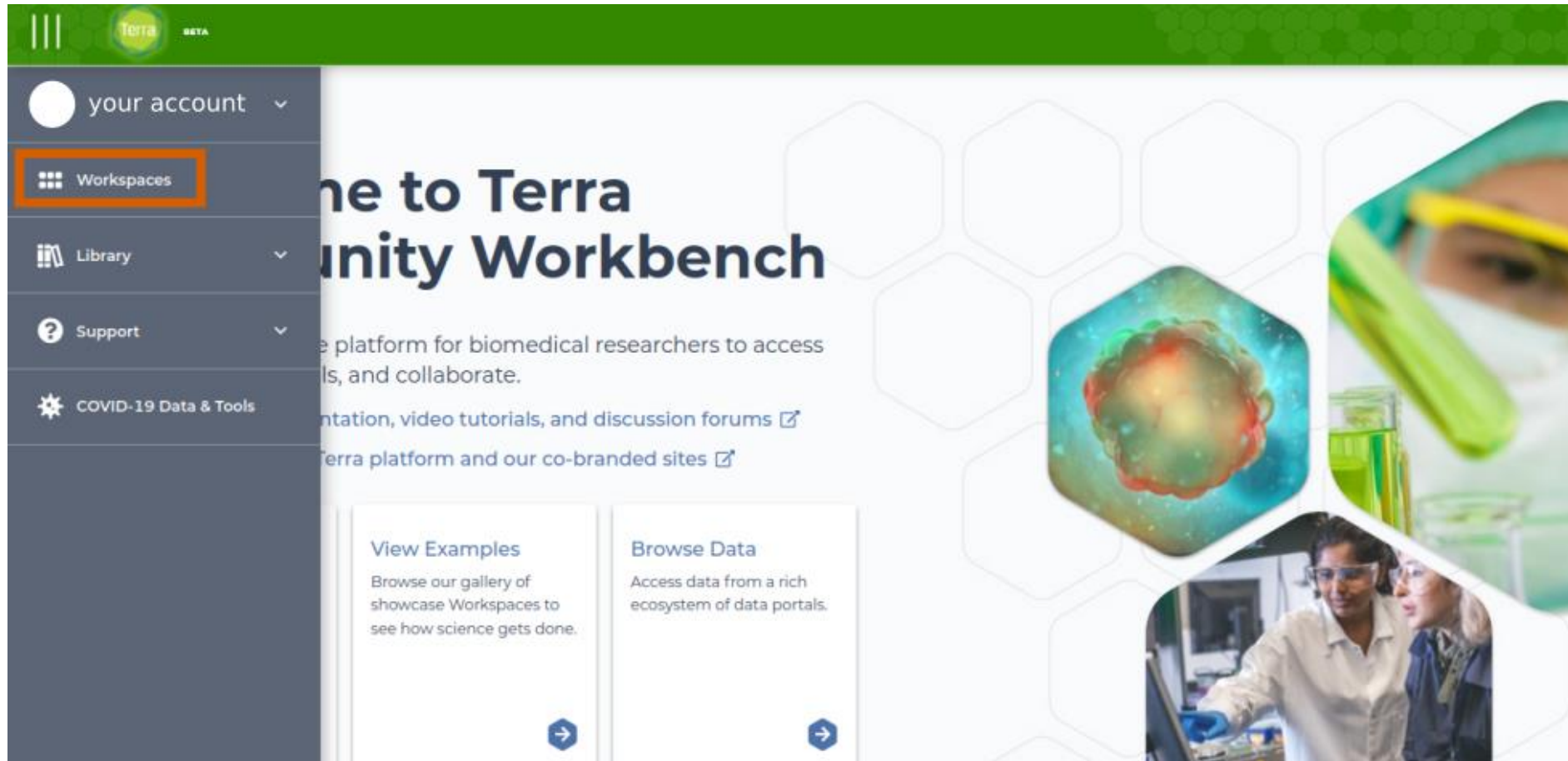
Terra uses cookies to enable the proper functioning and security of our website, and to improve your experience. the use of these functional cookies. If you do not wish to allow use of these cookies, you may tell us that by clicking here. To find out more, read our [privacy policy](#).

Give your google account email to Blythe

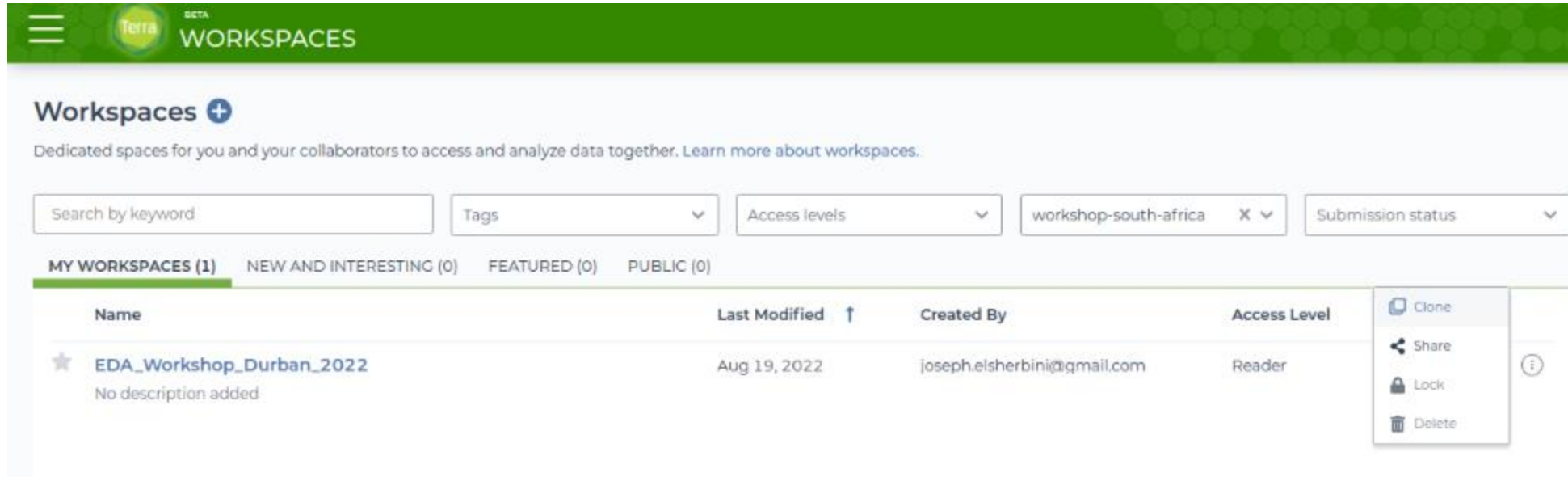
- [Please submit your Terra account email on this form](#) (link is on the logistics page of our workshop website). so we can manually add you to our Terra billing group for the workshop.
- If you are having issues, ask on Discord in the #terra-access channel



Click on the menu and click on workspaces



Click on the three dots to the right of EDA_Workshop_Durban_2022 and click Clone



The screenshot shows the Terra WORKSPACES interface. At the top, there's a green header with the Terra logo and the word "WORKSPACES". Below the header, there's a section titled "Workspaces +" with a subtitle "Dedicated spaces for you and your collaborators to access and analyze data together. [Learn more about workspaces.](#)".

Below the subtitle, there are several filters: "Search by keyword" (input field), "Tags" (dropdown), "Access levels" (dropdown), "workshop-south-africa" (tag with an 'x' to remove), and "Submission status" (dropdown).

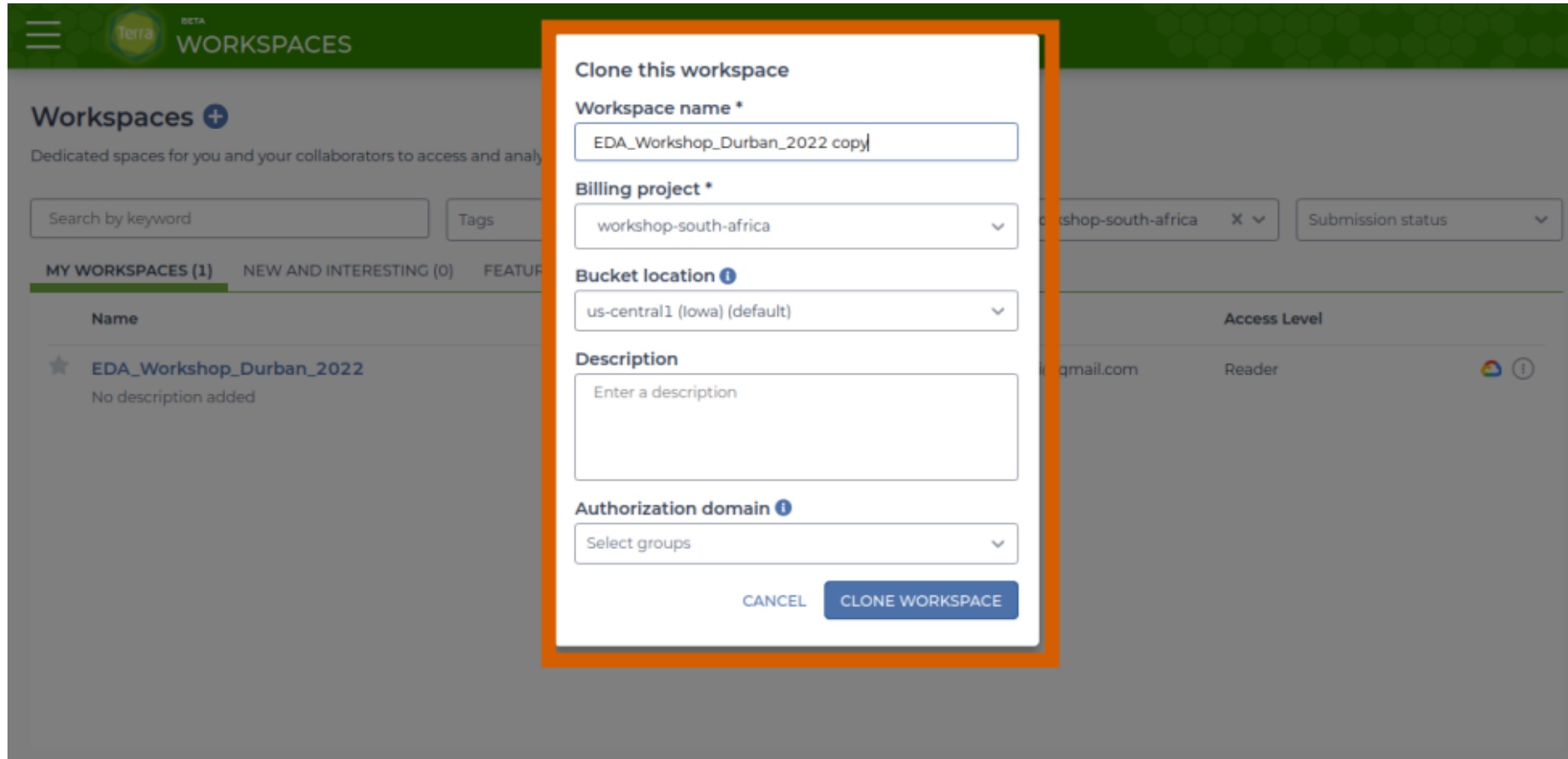
Under the filters, there are tabs: "MY WORKSPACES (1)", "NEW AND INTERESTING (0)", "FEATURED (0)", and "PUBLIC (0)". The "MY WORKSPACES (1)" tab is selected.

Below the tabs, there's a table with the following columns: "Name", "Last Modified", "Created By", and "Access Level".

Name	Last Modified	Created By	Access Level
★ EDA_Workshop_Durban_2022 No description added	Aug 19, 2022	joseph.elsherbini@gmail.com	Reader

To the right of the table, a context menu is open for the "EDA_Workshop_Durban_2022" workspace. The menu contains the following options: "Clone", "Share", "Lock", and "Delete".

You can name the workspace whatever you want, or leave it as is, and click "Clone Workspace"



The screenshot shows the Terraform Workspaces interface with a modal dialog for cloning a workspace. The dialog is titled "Clone this workspace" and contains the following fields:

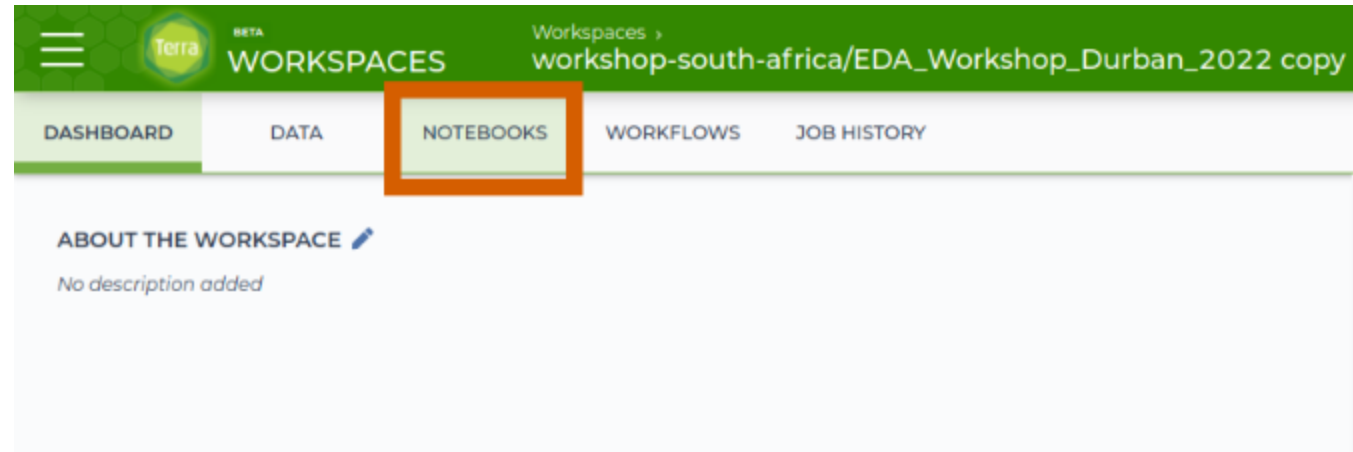
- Workspace name ***: A text input field containing "EDA_Workshop_Durban_2022 copy".
- Billing project ***: A dropdown menu with "workshop-south-africa" selected.
- Bucket location ⓘ**: A dropdown menu with "us-central1 (Iowa) (default)" selected.
- Description**: A text area with the placeholder "Enter a description".
- Authorization domain ⓘ**: A dropdown menu with "Select groups" selected.

At the bottom of the dialog are two buttons: "CANCEL" and "CLONE WORKSPACE". The background shows a list of workspaces, including "EDA_Workshop_Durban_2022".

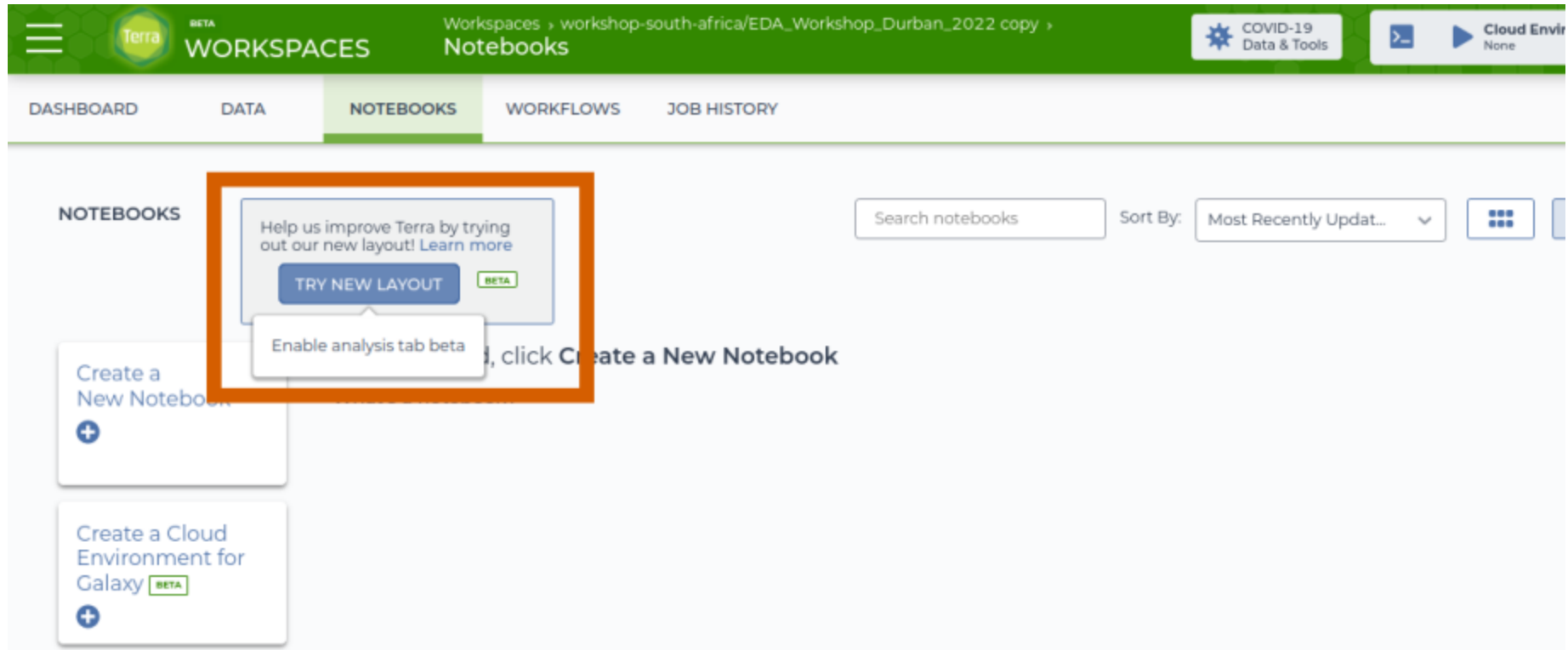
Billing Project should be workshop-south-africa, and you are welcome to add a description.

Authorization domain should be left as is

Click on NOTEBOOKS

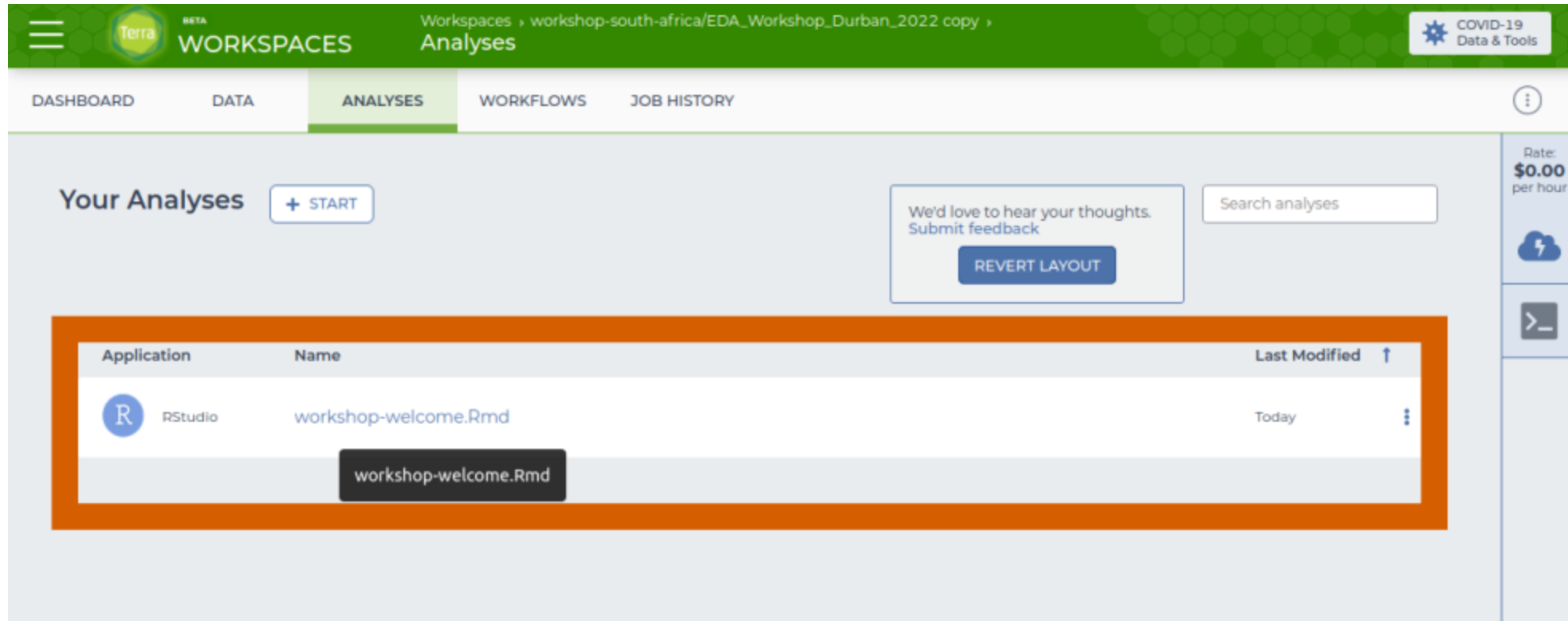


Click on TRY NEW LAYOUT



Skip the tour if it pops-up

Click on workshop-welcome.Rmd



The screenshot displays the Terra WORKSPACES interface. The top navigation bar is green and contains the Terra logo, the word "WORKSPACES", and a breadcrumb trail: "Workspaces > workshop-south-africa/EDA_Workshop_Durban_2022 copy > Analyses". A "COVID-19 Data & Tools" button is in the top right. Below the navigation bar is a tabbed interface with "DASHBOARD", "DATA", "ANALYSES" (selected), "WORKFLOWS", and "JOB HISTORY".

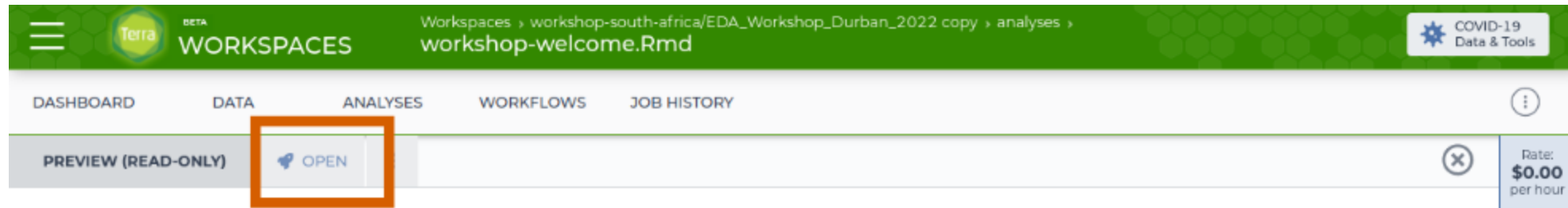
The main content area is titled "Your Analyses" and includes a "+ START" button. A feedback box with the text "We'd love to hear your thoughts. Submit feedback" and a "REVERT LAYOUT" button is also present. A search bar labeled "Search analyses" is located to the right of the feedback box.

A table titled "Your Analyses" is highlighted with an orange border. It has three columns: "Application", "Name", and "Last Modified". The table contains one row:

Application	Name	Last Modified
RStudio	workshop-welcome.Rmd	Today

Below the table, a dark button labeled "workshop-welcome.Rmd" is visible. On the right side of the interface, there is a sidebar with a "Rate: \$0.00 per hour" indicator, a lightning bolt icon, and a terminal icon.

Click on OPEN, and then don't click anything else until the next instructions!





Click on CUSTOMIZE on the bottom

RStudio Cloud Environment

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

Use default environment

CREATE

- RStudio (R 4.2.0, Bioconductor 3.15, Python 3.8.10) 
[What's installed on this environment?](#)
- Default compute size of **4 CPU(s)**, **15 GB memory**, and a **50 GB persistent disk** to keep your data even after you delete your compute
- [Learn more about Persistent disks and where your disk is mounted](#)
- This cloud environment will be created in the region **us-central1**. Copying data from a bucket in a different region may incur network egress charges. Note that network egress charges are not accounted for in cost estimates. For more information, particularly if you work with data stored in multiple cloud regions, please read the [documentation](#). 

Running cloud compute cost	Paused cloud compute cost	Persistent disk cost
\$0.21 per hr	\$0.01 per hr	\$2.00 per month

Create custom environment

CUSTOMIZE

Set your profile to 1 CPU

RStudio Cloud Environment ×

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

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

Application configuration ⓘ

RStudio (R 4.2.0, Bioconductor 3.15, Python 3.8.10) ▼

What's installed on this environment? Updated: Jun 26, 2022
Version: 3.15.2 ⓘ

Cloud compute profile

CPUs 1 ▼ **Memory (GB)** 3.75 ▼

☐ **Enable GPU** 1 ✓ TA [Learn more about GPU cost and restrictions.](#) ⓘ

Startup 2

URI 4

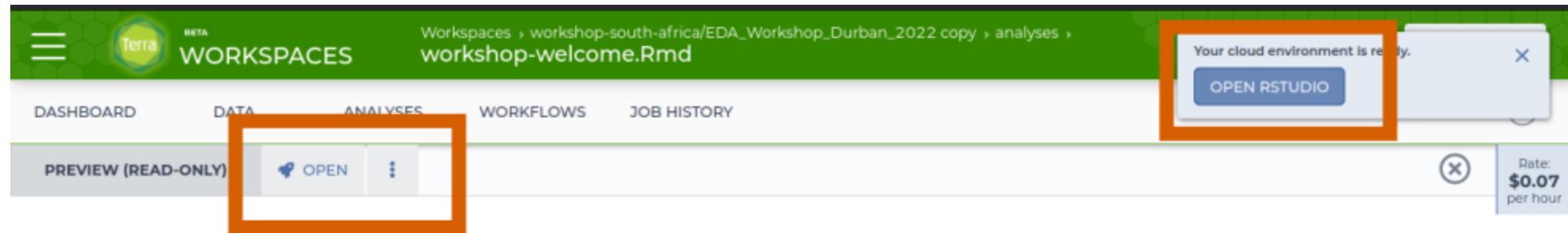
Compute 8

Standard 16 ▼

32

Click CREATE at the bottom of the window
(you may need to scroll down to see it)

Setting up for the first time takes 5 minutes
After it's ready, click OPEN



Note: each day you'll have to click OPEN, but it'll only take 1-2 minutes to start after the first time

Now you should have RStudio running on Terra!

