Intro to Terra

February, 2022

Table of Contents

# About this Course

# 1 Introduction

The health and life sciences data revolution has arrived. Accurate and inexpensive genome sequencing is a reality, high-resolution imaging is becoming routine, and clinical data is increasingly available in machine-readable formats. These breakthroughs bring us to the threshold of a new era in biomedicine, one where the data sciences hold the potential to propel our understanding and treatment of human disease.

Enter [Terra](http://terra.bio), a scalable platform for biomedical research. Terra’s mission is to enable the next generation of collaborative biomedical research by building an open platform that connects researchers to each other and to the datasets and tools they need to achieve scientific breakthroughs. But building infrastructure is only the first step. We are collaborating with funders, data generators, method developers and community leaders to support and empower researchers in key domains as they transition their work to this new ecosystem. This includes the Analysis, Visualization and Informatics Labspace (AnVIL) funded by the National Human Genome Research Institute, BioData Catalyst funded by the National Heart, Lung and Blood Institute, and FireCloud funded by the National Cancer Institute.

Researchers are already leveraging Terra to accelerate their research. Cancer genomics, infectious disease, medical and population genetics, and single-cell transcriptomics are just some of the biomedical disciplines supported by Terra.

This course is meant to provide a guided learning experience that gives you a foundational understanding of Terra and the key concepts underpinning its principles and intended applications. You will be working with the “generic” Terra application, but if you were working in the context of a project powered by Terra (i.e. [AnVIL](https://anvil.terra.bio/), [BioData Catalyst](https://terra.biodatacatalyst.nhlbi.nih.gov/), [FireCloud](https://firecloud.terra.bio/)), your experience in the application will be the same even if the “branding” looks a little different.

By the end of this course, you will be able to:

1. Describe Terra’s goals and its guiding principles,
2. Name the types of analysis that you can do in Terra,
3. Define key cloud computing components that will enable you to work in Terra,
4. Securely access Terra with your own account,
5. Access Terra support to learn more and get help when you need it,
6. Articulate your next steps to getting started with Terra.

## 1.1 Course materials

Most of the reading you will do in this course will be from articles in the [Terra Support Center](http://support.terra.bio). You can think of that knowledge base as the textbook to this course.

We use this approach because Terra is in active development, with features continually being improved and released. The knowledge base will be more rapidly updated than this course, and so the reading assignments remain as up-to-date as possible.

We also assign a small number of videos that are hosted on the [Terra YouTube channel](https://www.youtube.com/c/TerraBioApp/videos). We don’t currently provide transcripts of the videos, but you can make use of the closed caption feature available in YouTube if desired.

In some places, we suggest additional optional reading. While not required for the course, we want to make sure you are aware of these resources if you are ready to take a deeper dive on your own.

## 1.2 Quizzes

There are two quizzes in this course to test your understanding of Terra before you move on.

## 1.3 Course discussion

We welcome your feedback and encourage you to discuss the course with other learners in our [Community Forum](https://support.terra.bio/hc/en-us/community/topics). (You’ll learn more about the forum in a later section). %% I’ll probably want to link directly to a dedicated section of the forum. Just link to the forum itself for now.

## 1.4 Acknowledgements

This course is authored by the [Data Sciences Platform](https://www.broadinstitute.org/data-sciences-platform) at the [Broad Institute of MIT and Harvard](https://www.broadinstitute.org/).

Terra is developed by the Broad Institute of MIT and Harvard in collaboration with [Microsoft](https://www.microsoft.com/en-us/industry/health/pharmaceuticals) and [Verily](https://verily.com/our-story/about-us/).

# 2 Terra overview

Let’s begin with a short introductory video about Terra. After watching this video, you should understand Terra’s mission and be familiar with its core functionality.

You’ll see that the video is presented as a “video assignment.” (And you’ll similarly see reading assignments just a little later). To move smoothly through the course, watch or read each assigned resource before moving on. We should point out that although our articles routinely link readers to further resources and more detailed articles for those who want to learn more, we don’t expect that you will do so in this course. Reading just the assigned article is sufficient. (Though if you are curious and want to keep reading, please do!).

Video assignment (2 minutes 39 seconds):

[Terra in a nutshell](https://youtu.be/w9mGcvU65r0)

After viewing the video, ask yourself: - What is the advantage of a cloud-centric approach to data sharing? - How does Terra help facilitate data analysis for biomedical researchers? - For which biomedical research communities does Terra serve as a hub?

Next is some reading to supplement what you just watched. The following article describes more about Terra’s functionality, particularly the types of analysis you can do – without going too deep into configurations and other details just yet.

Reading assignment (7 minute read): [Biomedical research in the cloud on Terra](https://support.terra.bio/hc/en-us/articles/360022714931)

While reading the article, ask yourself: - What kinds of analyses can I do in Terra? - Which types of analysis are suited for pipelining with workflows? - Which interactive analysis tools can I run in Terra?

## 2.1 New to cloud

If you are new to the cloud entirely, making the transition from your local computing environment to Terra might seem daunting. But it doesn’t need to be. For this introductory course we won’t take you on a deep-dive into cloud computing, but you will need to be aware of some core concepts. This will help you understand the architecture of Terra in the section following this one.

Reading assignment (3 minute read): [New to Cloud](https://terra.bio/resources/new-to-cloud/)

While reading this brief overview on the Terra website, ask yourself: - Where are files stored in the cloud? - How are computing resources and software organized in the cloud? - How do I pay for cloud computing expenses? - How is access to resources and spending controlled? - What are some important considerations for storing data in the cloud?

## 2.2 Terra and data in the cloud

After that high-level introduction to the cloud in the previous section, let’s now look more into Terra’s relationship with data, and how it is different from working in an institutional high-performance computing environment or other local analysis platform. For example, where exactly is data stored? How do you access, organize, and interact with data in Terra? How do you find data libraries from which to import data to Terra? In the assigned reading, you’ll learn how cloud storage and the Terra Cloud Environment integrate with the Terra workspace and how your files live across different resources.

Reading assignment (7 minute read): [Terra architecture and how your files live in it](https://support.terra.bio/hc/en-us/articles/360058163311)

While reading the article, ask yourself: - What are the basic components of Terra? - Where can data be stored? - What are the basic components of the Terra Cloud Environment? - How do cloud resources communicate so that I can move files between them?

If there are any cloud terminology terms you come across that you aren’t familiar with, you can refer to our [glossary of terms related to cloud-based genomics](https://support.terra.bio/hc/en-us/articles/4402431295245).

# About the Authors

These credits are based on our [course contributors table guidelines](https://github.com/jhudsl/OTTR_Template/wiki/How-to-give-credits).

| Credits | Names |
| --- | --- |
| **Pedagogy** |  |
| Lead Content Instructor(s) | [FirstName LastName](link%20to%20personal%20website) |
| Lecturer(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | Delivered the course in some way - video or audio |
| Content Author(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | If any other authors besides lead instructor |
| Content Contributor(s) (include section name/link in parentheses) - make new line if more than one section involved | Wrote less than a chapter |
| Content Editor(s)/Reviewer(s) | Checked your content |
| Content Director(s) | Helped guide the content direction |
| Content Consultants (include chapter name/link in parentheses or word “General”) - make new line if more than one chapter involved | Gave high level advice on content |
| Acknowledgments | Gave small assistance to content but not to the level of consulting |
| **Production** |  |
| Content Publisher(s) | Helped with publishing platform |
| Content Publishing Reviewer(s) | Reviewed overall content and aesthetics on publishing platform |
| **Technical** |  |
| Course Publishing Engineer(s) | Helped with the code for the technical aspects related to the specific course generation |
| Template Publishing Engineers | [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/) |
| Publishing Maintenance Engineer | [Candace Savonen](https://www.cansavvy.com/) |
| Technical Publishing Stylists | [Carrie Wright](https://carriewright11.github.io/), [Candace Savonen](https://www.cansavvy.com/) |
| Package Developers ([ottr](https://github.com/jhudsl/ottr)) | [John Muschelli](https://johnmuschelli.com/), [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/) |
| **Art and Design** |  |
| Illustrator(s) | Created graphics for the course |
| Figure Artist(s) | Created figures/plots for course |
| Videographer(s) | Filmed videos |
| Videography Editor(s) | Edited film |
| Audiographer(s) | Recorded audio |
| Audiography Editor(s) | Edited audio recordings |
| **Funding** |  |
| Funder(s) | Institution/individual who funded course including grant number |
| Funding Staff | Staff members who help with funding |

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.3 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2022-02-16   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.3)   
## bookdown 0.24 2022-02-15 [1] Github (rstudio/bookdown@88bc4ea)   
## callr 3.4.4 2020-09-07 [1] RSPM (R 4.0.2)   
## cli 2.0.2 2020-02-28 [1] RSPM (R 4.0.0)   
## crayon 1.3.4 2017-09-16 [1] RSPM (R 4.0.0)   
## desc 1.2.0 2018-05-01 [1] RSPM (R 4.0.3)   
## devtools 2.3.2 2020-09-18 [1] RSPM (R 4.0.3)   
## digest 0.6.25 2020-02-23 [1] RSPM (R 4.0.0)   
## ellipsis 0.3.1 2020-05-15 [1] RSPM (R 4.0.3)   
## evaluate 0.14 2019-05-28 [1] RSPM (R 4.0.3)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.6.1 2022-01-22 [1] CRAN (R 4.0.2)   
## htmltools 0.5.0 2020-06-16 [1] RSPM (R 4.0.1)   
## knitr 1.33 2022-02-15 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.0 2021-02-15 [1] CRAN (R 4.0.2)   
## magrittr 2.0.2 2022-01-26 [1] CRAN (R 4.0.2)   
## memoise 1.1.0 2017-04-21 [1] RSPM (R 4.0.0)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgload 1.1.0 2020-05-29 [1] RSPM (R 4.0.3)   
## prettyunits 1.1.1 2020-01-24 [1] RSPM (R 4.0.3)   
## processx 3.4.4 2020-09-03 [1] RSPM (R 4.0.2)   
## ps 1.3.4 2020-08-11 [1] RSPM (R 4.0.2)   
## purrr 0.3.4 2020-04-17 [1] RSPM (R 4.0.3)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 0.4.10 2022-02-15 [1] Github (r-lib/rlang@f0c9be5)   
## rmarkdown 2.10 2022-02-15 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.2 2020-11-15 [1] CRAN (R 4.0.2)   
## sessioninfo 1.1.1 2018-11-05 [1] RSPM (R 4.0.3)   
## stringi 1.5.3 2020-09-09 [1] RSPM (R 4.0.3)   
## stringr 1.4.0 2019-02-10 [1] RSPM (R 4.0.3)   
## testthat 3.0.1 2022-02-15 [1] Github (R-lib/testthat@e99155a)   
## usethis 2.1.5.9000 2022-02-15 [1] Github (r-lib/usethis@57b109a)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2022-02-15 [1] Github (yihui/xfun@74c2a66)   
## yaml 2.2.1 2020-02-01 [1] RSPM (R 4.0.3)   
##   
## [1] /usr/local/lib/R/site-library  
## [2] /usr/local/lib/R/library

# References