Bioinformatics in the Cloud with the CGC

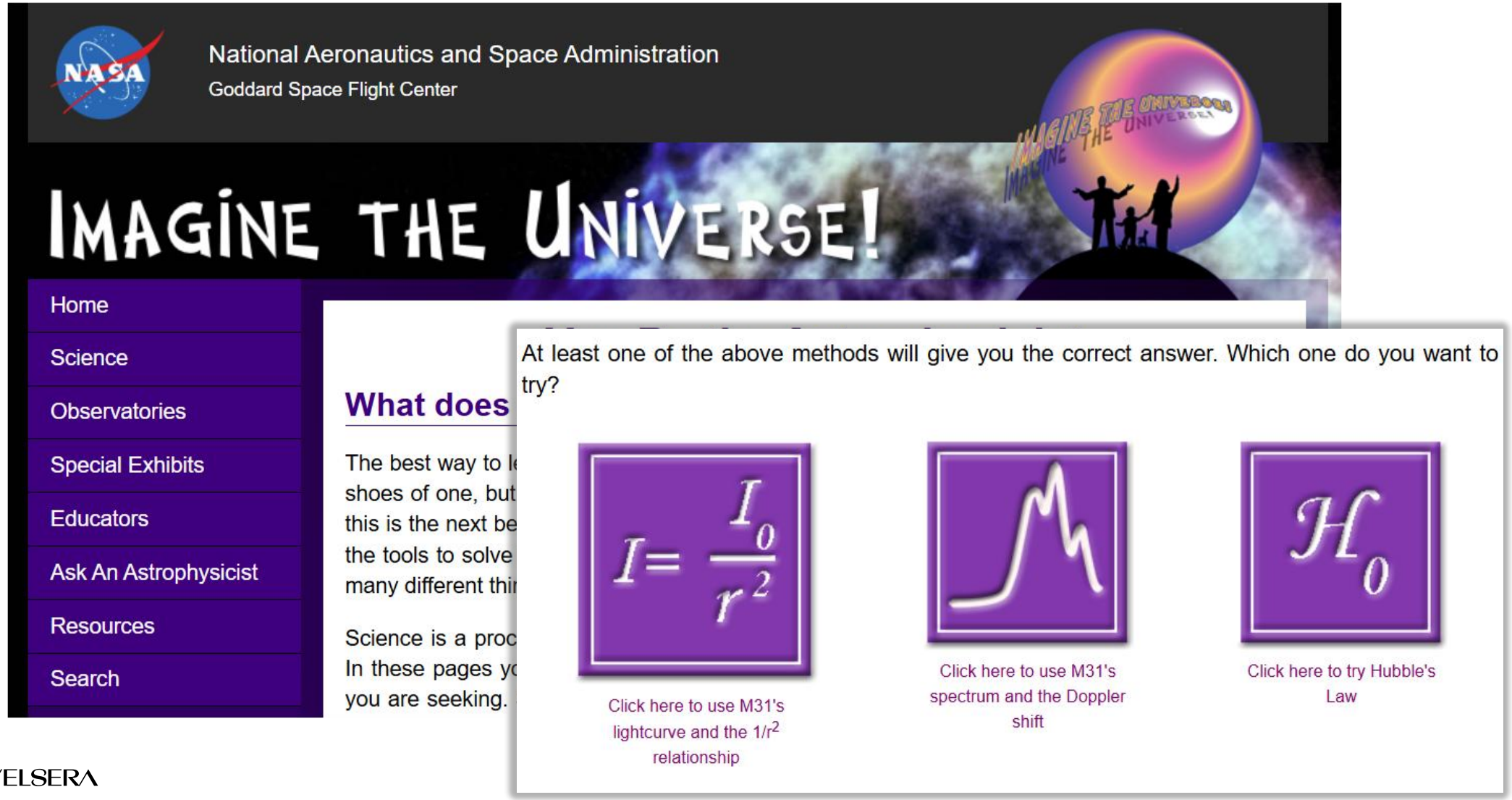
Leveraging the cloud to turbocharge your research

BTEP webinar

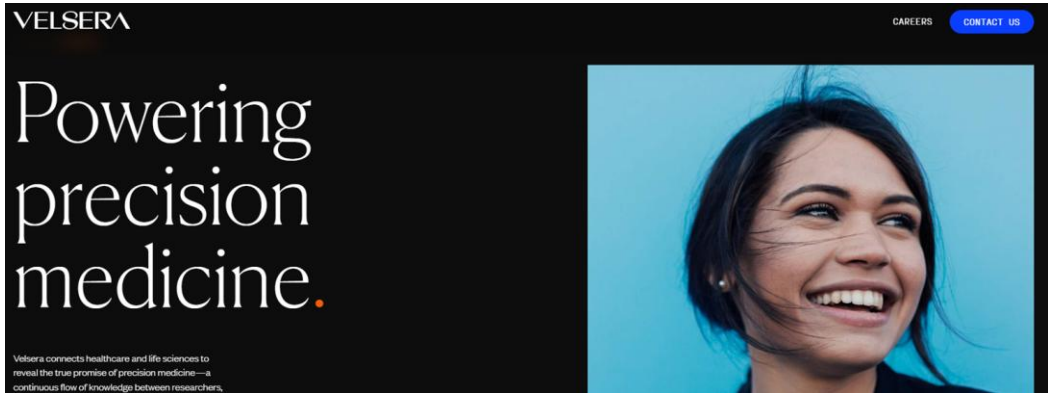
07/21/2026

Jeni Cliffe

Mini-bio – Jenepher A. Cliffe (M.S. Astrophysics)



Community Engagement Manager, Velsera



- Make NIH research better by:
  - Helping scientists and doctors analyze their research data on the Cloud
  - Teaching people how to work with and manage big data
  - Giving people better tools so research happens faster

From exploding stars to – humans?!

Agenda.

Bioinformatics in the Cloud with CGC

01	Introduction to the CGC and Cloud Computing
02	Accessing data and running tools (overview)
03	Variant analysis run-through
04	New feature highlights: OHIF viewer and 3D Slicer

Introduction to Cloud Computing with the Cancer Genomics Cloud

Let’s take a poll!

How comfortable are you with the cloud and cloud data?

1

Not comfortable

I’ve used it very little, or never

2

Somewhat comfortable

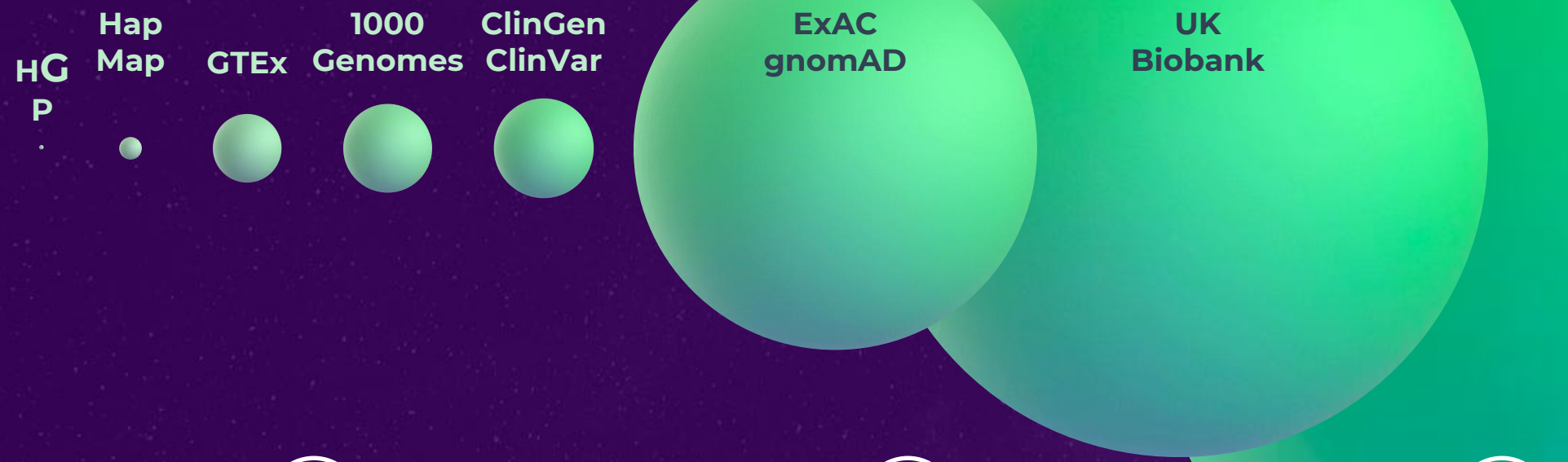
I’ve used it before

3

Very comfortable

I use it all the time

Data generation is on a roll...



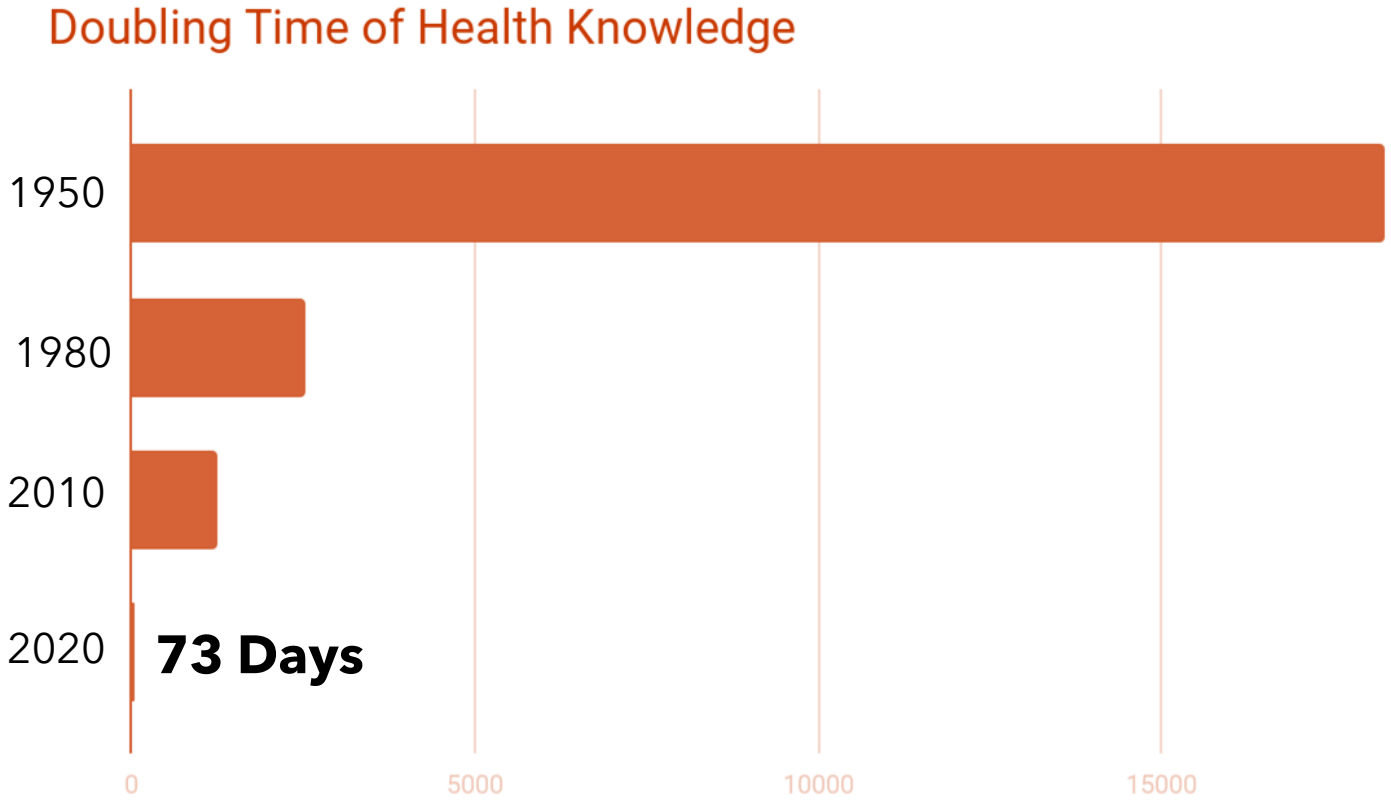
2001
1
Human Genome

2010
1000
Human Genomes

2020
1M
Human Genomes

VELSERA

The Era of Big Data in Health

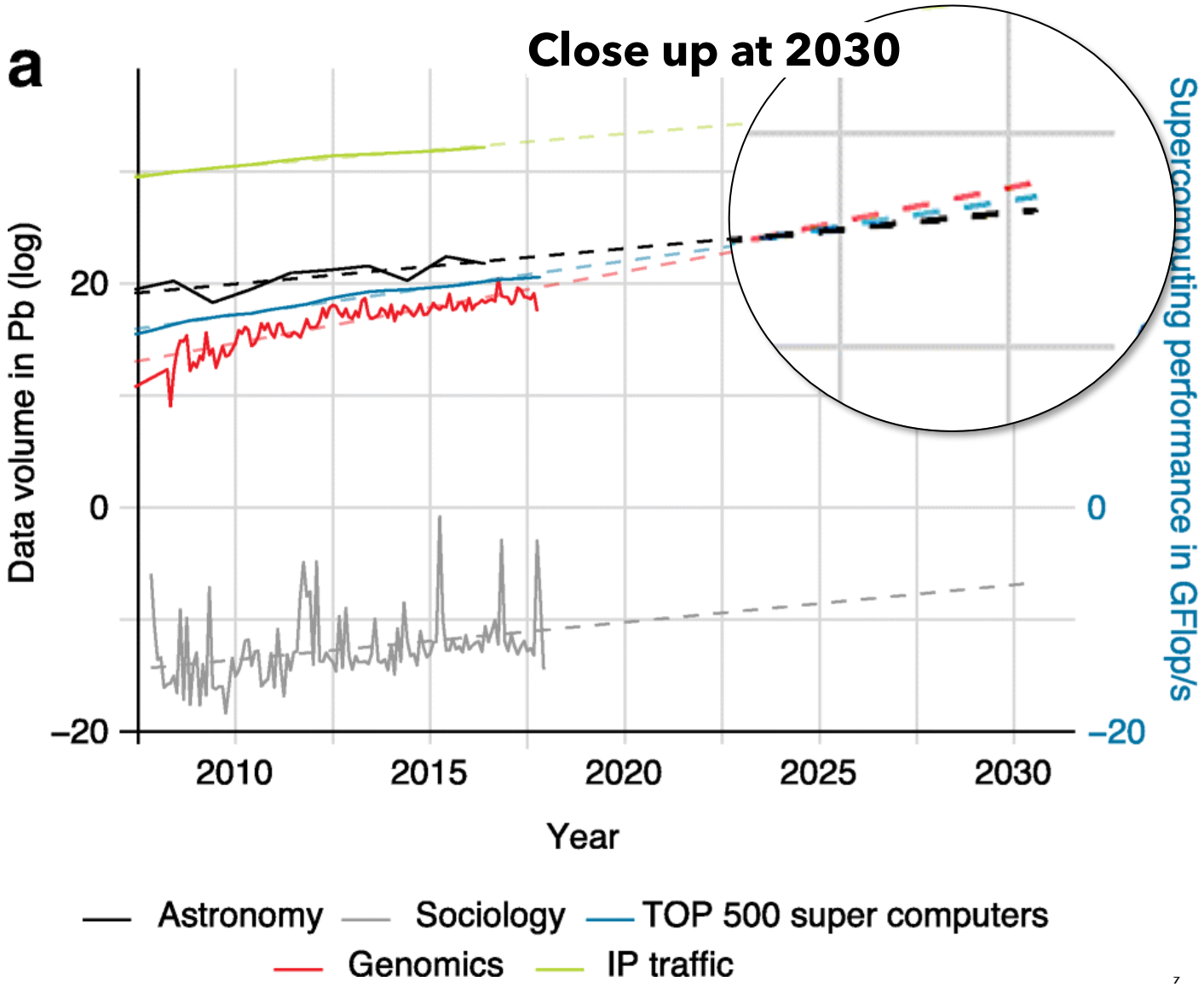


More biomedical data were generated in 2021 than all previous years **combined**

More different ***kinds*** of data – gene and genome sequencing, imaging, sensors (i.e. Fitbit), cellular measurement data, and more

How much data is it?

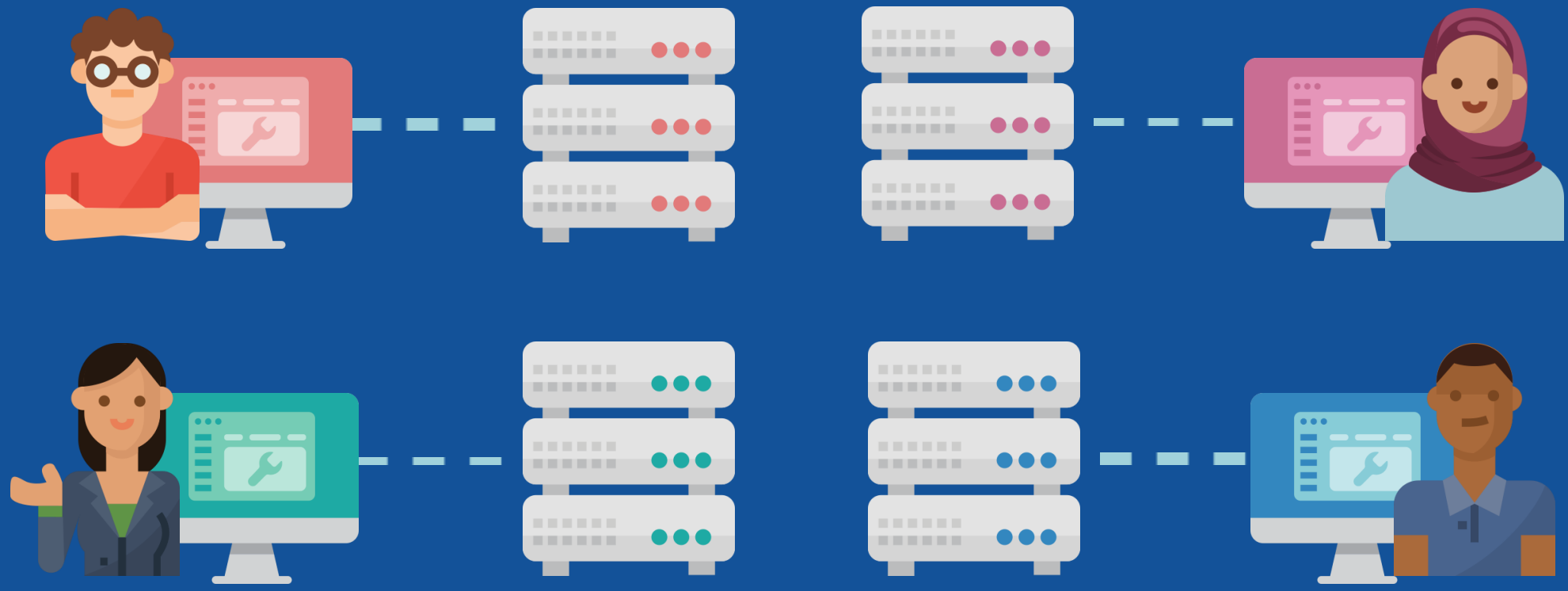
- *Estimated* between 2 and 40 exabytes of new data per year (Stephens *et al.* 2015)
- Genomics data is growing **faster than *the internet***
- Analysis-to-Data Paradigm – analyze data "where it lives"



How do we share and analyze datasets?

Traditional Approach

Bring data to researchers



VELSERA

CANCER GENOMICS CLOUD
SEVEN BRIDGES

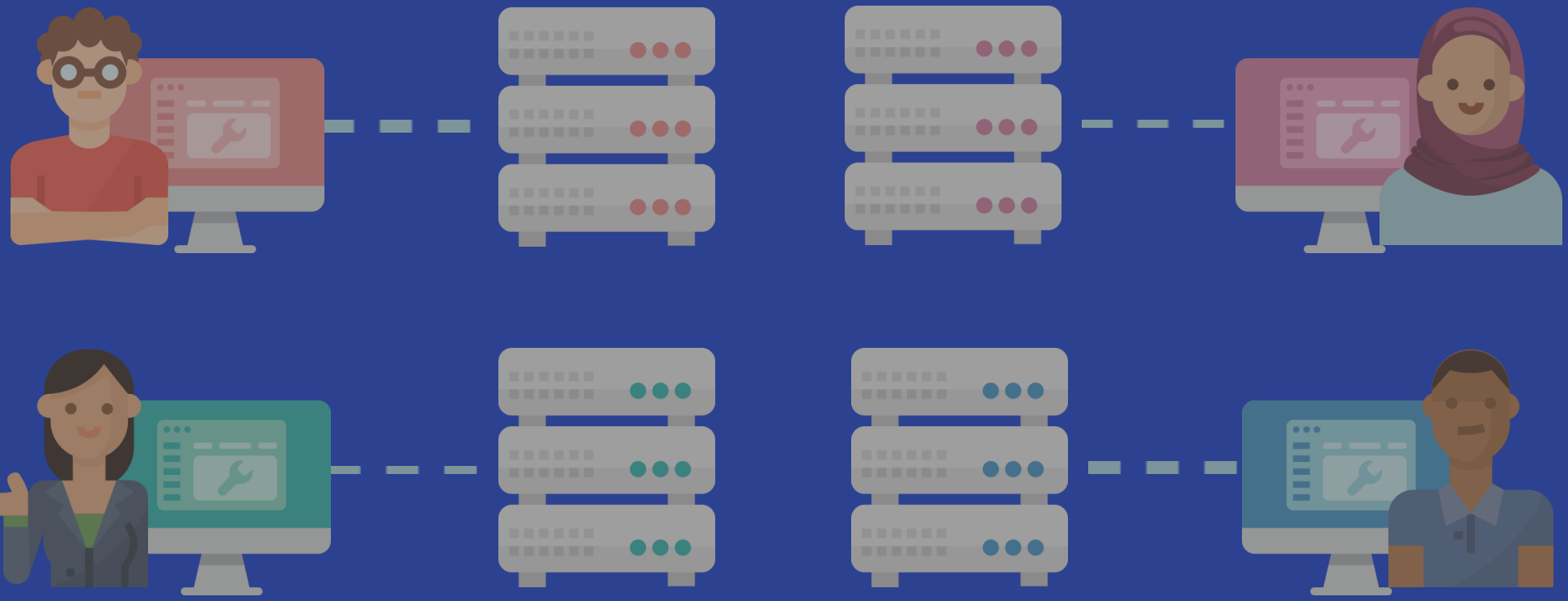
What’s wrong with this traditional approach? ?

Data is copied, not shared

High storage & infrastructure costs

Hard to share/ reproduce

Security on individuals



A new cloud-native vision for biomedical research

VELSERA

CANCER GENOMICS CLOUD
SEVEN BRIDGES

A new cloud-native vision for biomedical research

~~Data is copied, not shared~~

True immediate data sharing

~~High storage & infrastructure costs~~

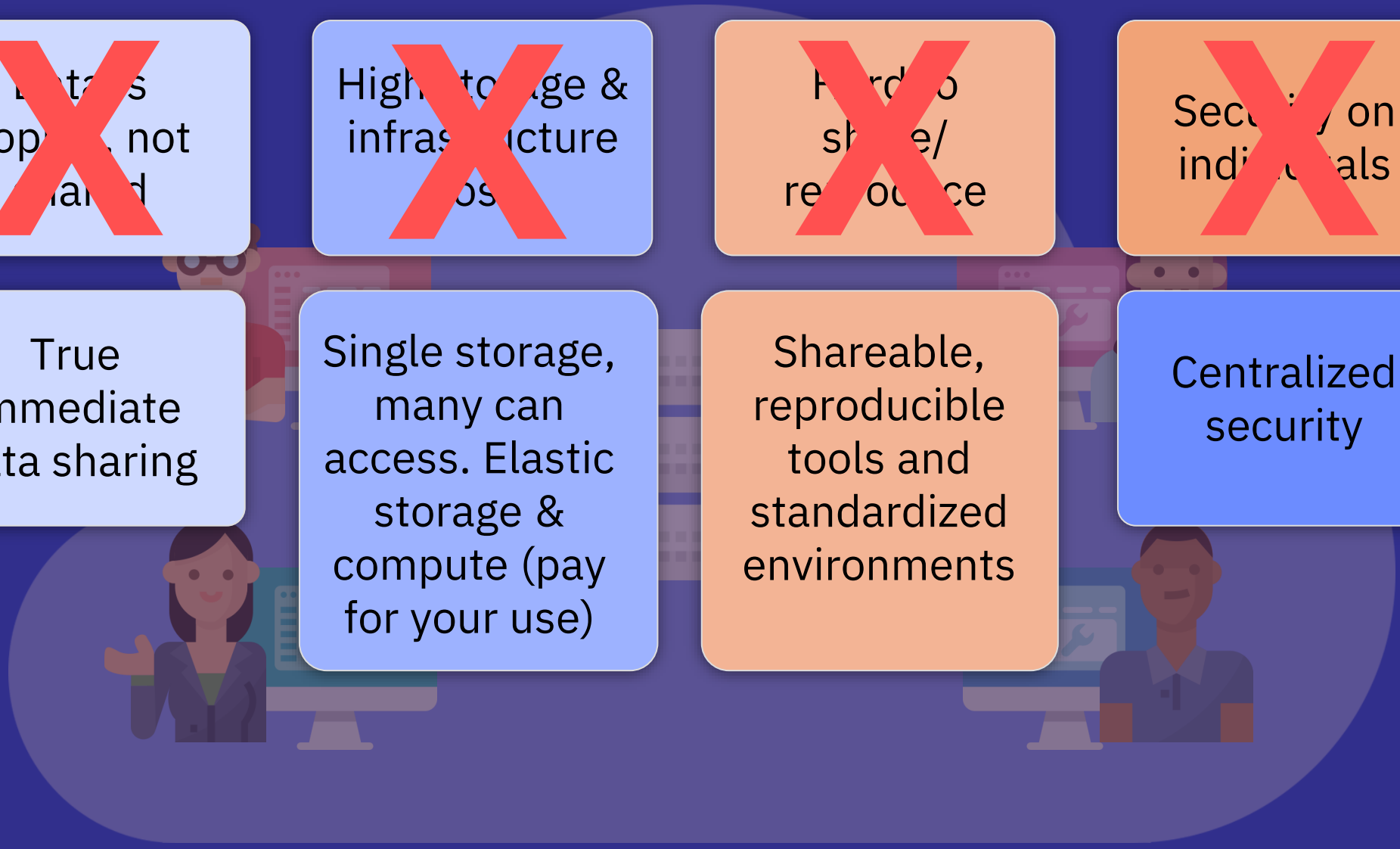
Single storage, many can access. Elastic storage & compute (pay for your use)

~~Hard to share/ reproduce~~

Shareable, reproducible tools and standardized environments

~~Security on individuals~~

Centralized security

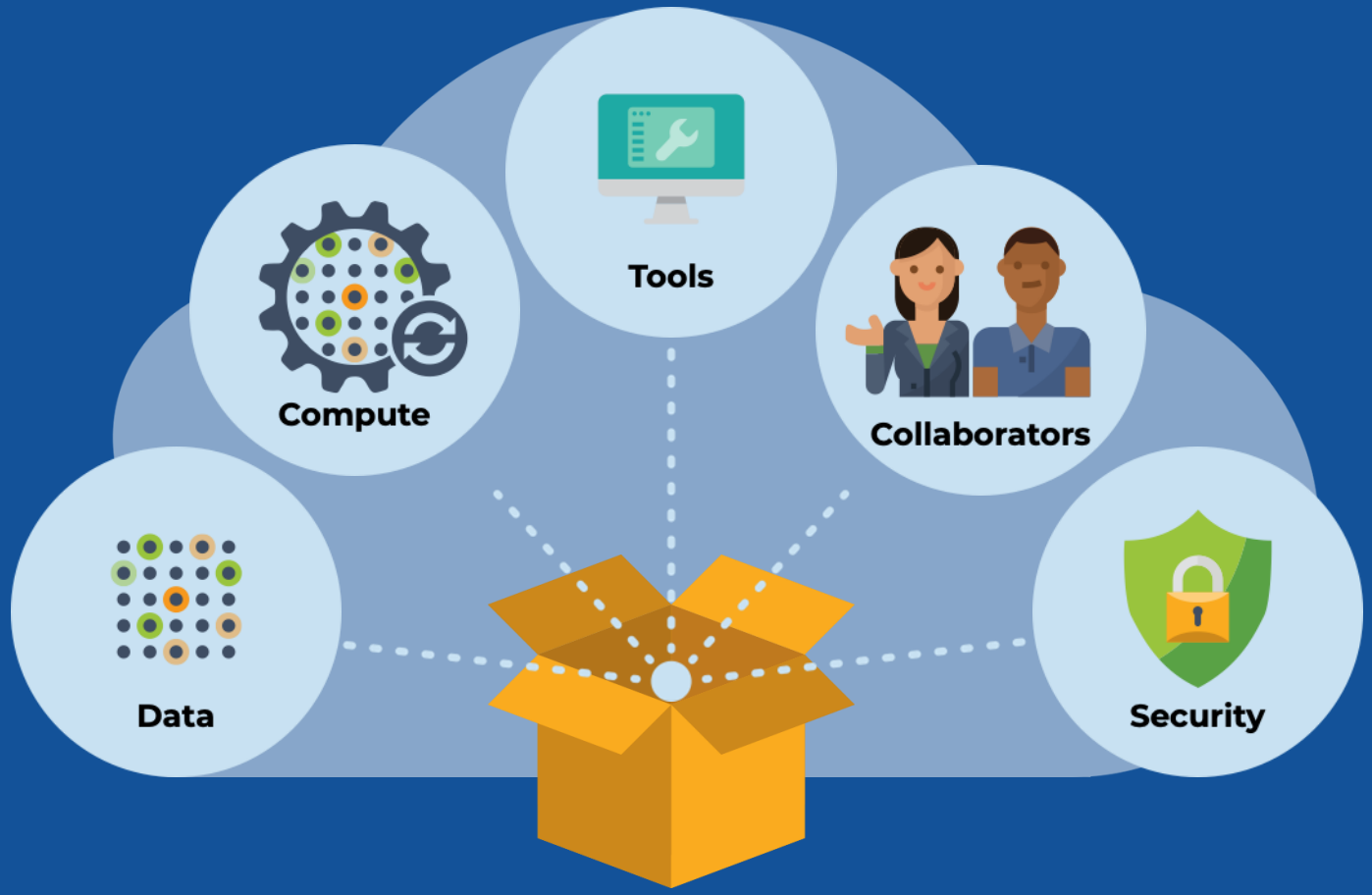


VELSERA

CANCER GENOMICS CLOUD
SEVEN BRIDGES

A new cloud-native vision for bioinformatics

Migrating to the cloud can be challenging!!



VELSERA

CANCER GENOMICS CLOUD
SEVEN BRIDGES

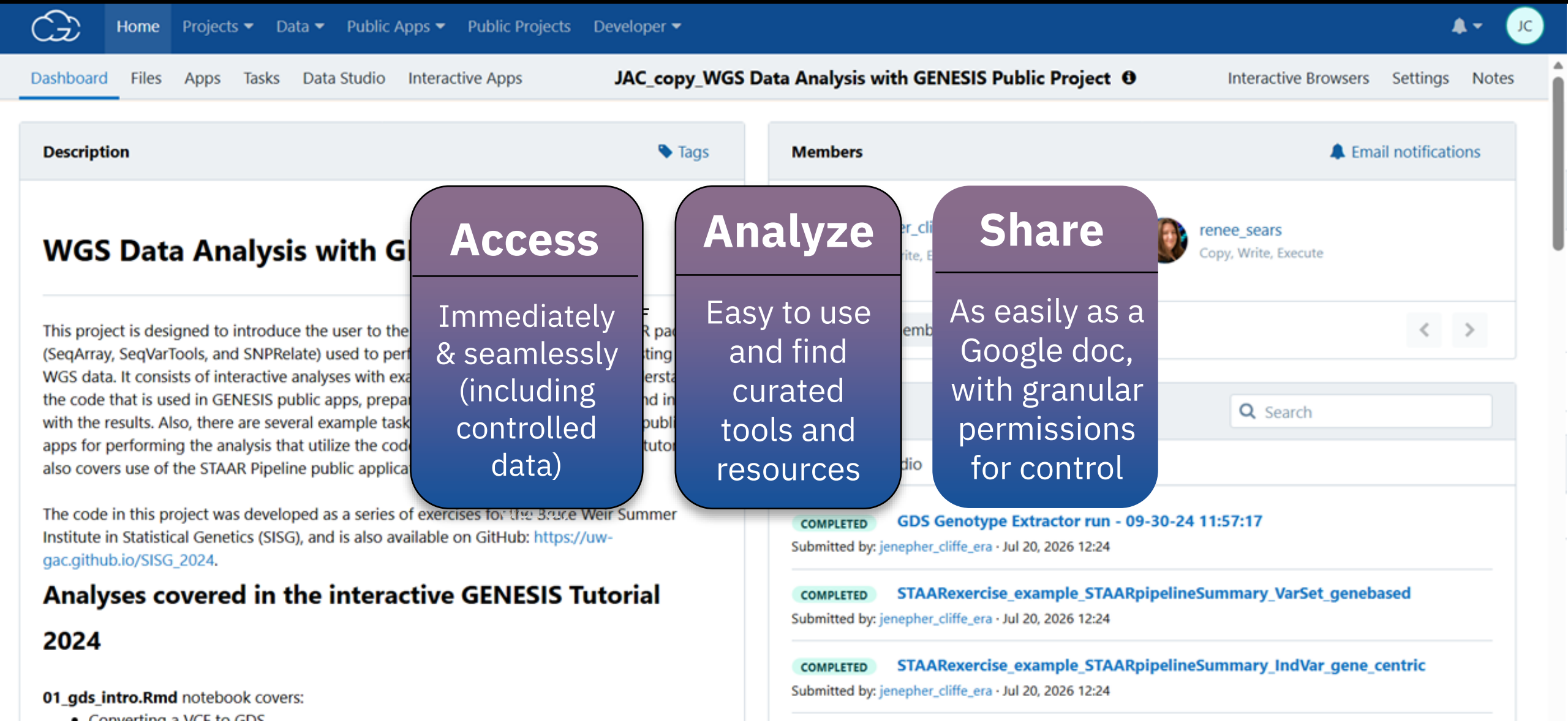
CANCER GENOMICS CLOUD

SEVEN BRIDGES

cancergenomicscloud.org

VELSERA

A new cloud-native vision for biomedical research



Access

Immediately & seamlessly (including controlled data)

Analyze

Easy to use and find curated tools and resources

Share

As easily as a Google doc, with granular permissions for control

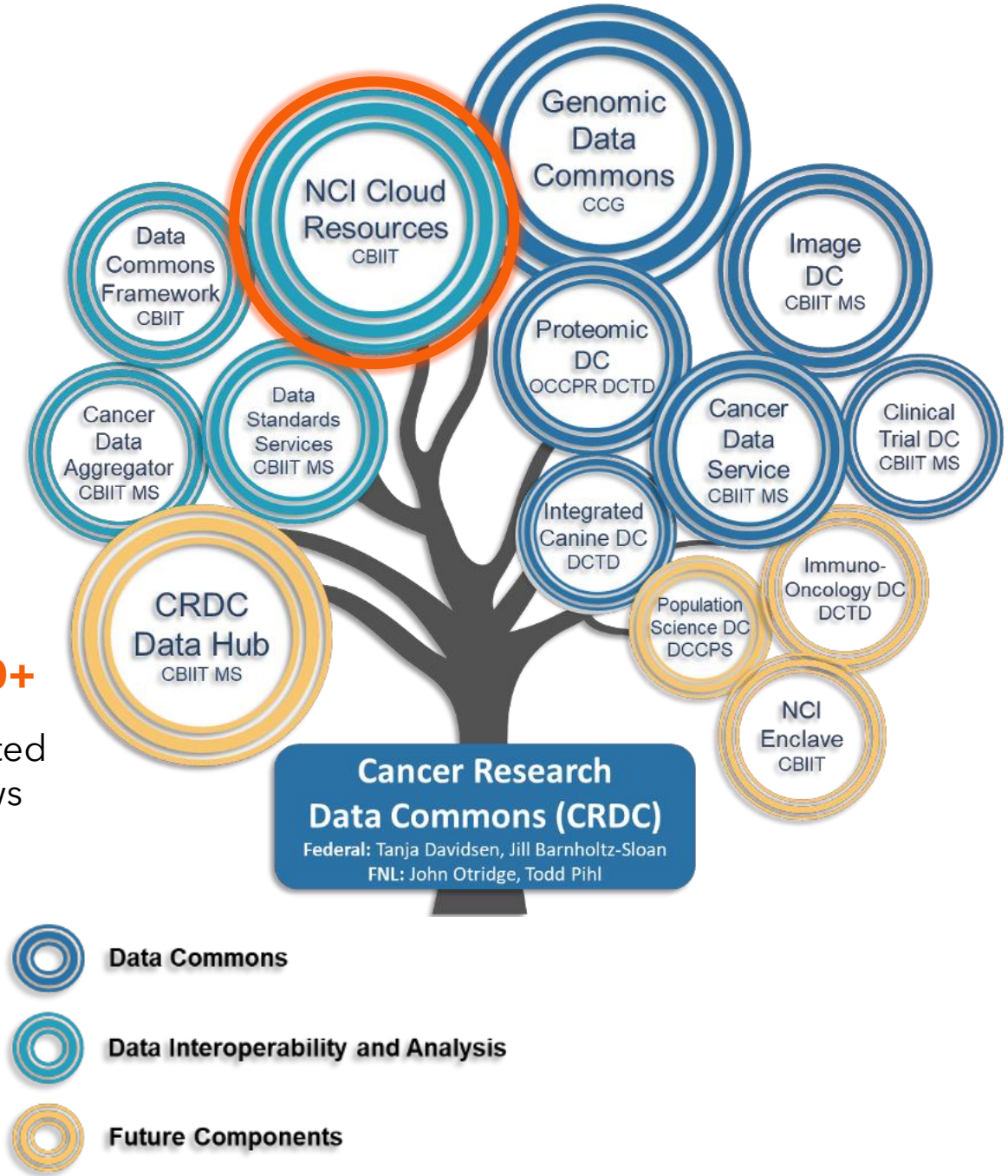
VELSERA

CANCER GENOMICS CLOUD
SEVEN BRIDGES

Powerful, yet easy to use interfaces to empower cancer researchers to draw new insights from petabyte scale data.

Stable, secure, and highly customizable cloud storage and computing platform

3+	2.4M+	1300+	12,500+	109,000+
Petabytes Public Data	Years of Compute	Public Tools & Workflows	Users	User-Created Workflows



VELSERA

Who are CGC Users?

The CGC is designed to serve a wide range of scientists and users with varying skill sets

ADMINISTRATORS

- Manage and control users
- Monitor and control institutional assets
- Manage and monitor projects
- Monitor and control costs
- Create reports

BIOINFORMATICIANS

- Store, manage, and share data
- Access public and proprietary datasets
- Query, build, and investigate cohorts of interest
- Access optimized tools and workflows
- Create, optimize, maintain, and distribute new tools and workflows
- Create bush-button automation solutions
- Analyze data at scale with tools and workflows
- Conduct interactive exploratory analyses
- Explore/visualize results/gather insights
- Easily collaborate with stakeholders
- Integrate with external systems

BENCH SCIENTISTS

- Store, manage, and share data
- Run optimized tools/ workflows at scale
- Conduct defined analyses via push-button solutions
- Investigate/visualize results
- Easily collaborate with other stakeholders

CLINICIANS

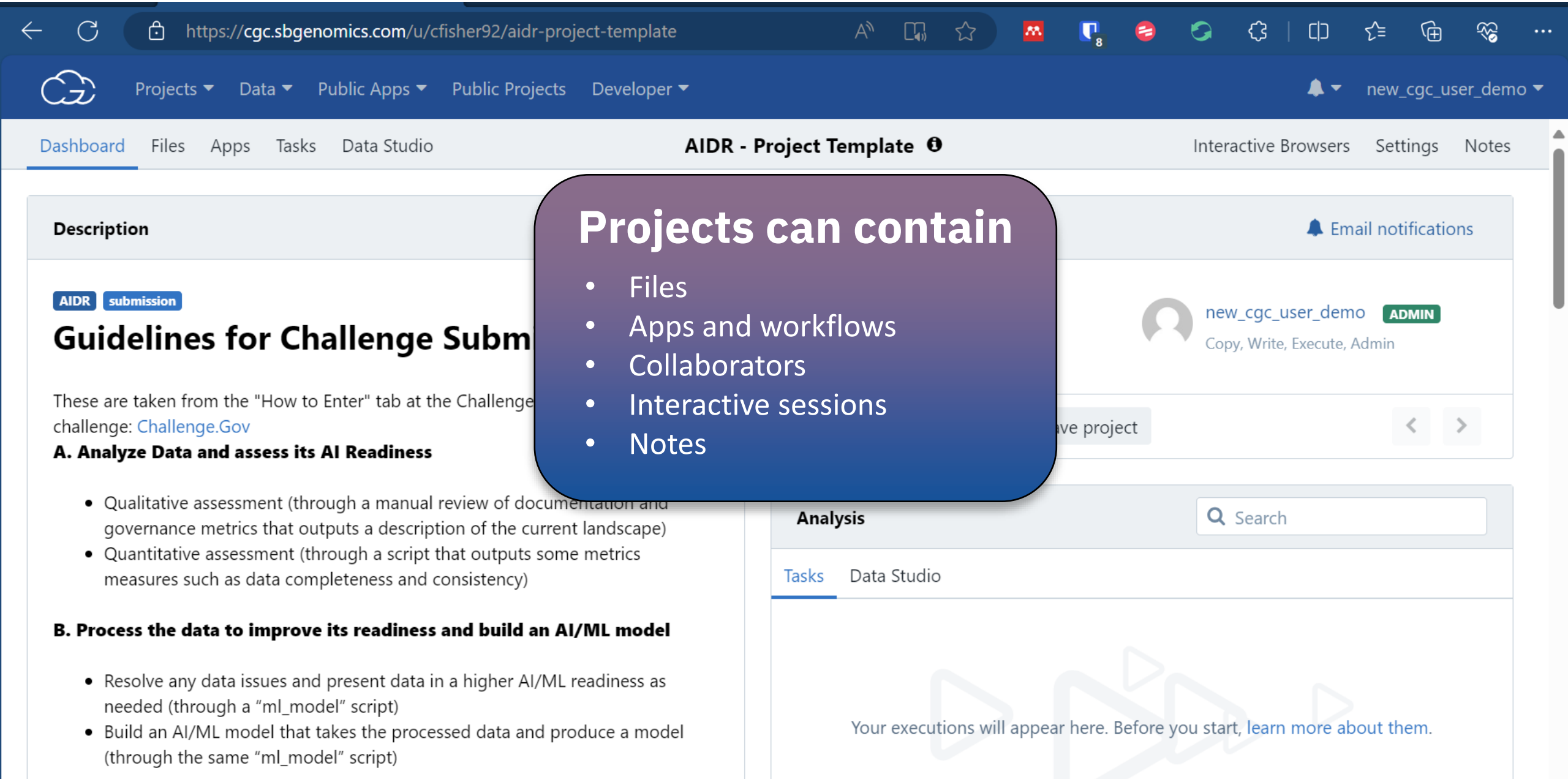
- Conduct validated analyses via push-button solutions
- Query, build, and investigate cohorts of interest
- Create reports
- Investigate/visualize results
- Easily collaborate with other stakeholders

DEVELOPERS

- Create, optimize, and maintain new tools and workflows
- Create push-button automation solutions
- Create custom interfaces for specific use cases
- Distribute proprietary tools/workflows
- Integrate upstream/downstream

Navigating the Cancer Genomics Cloud

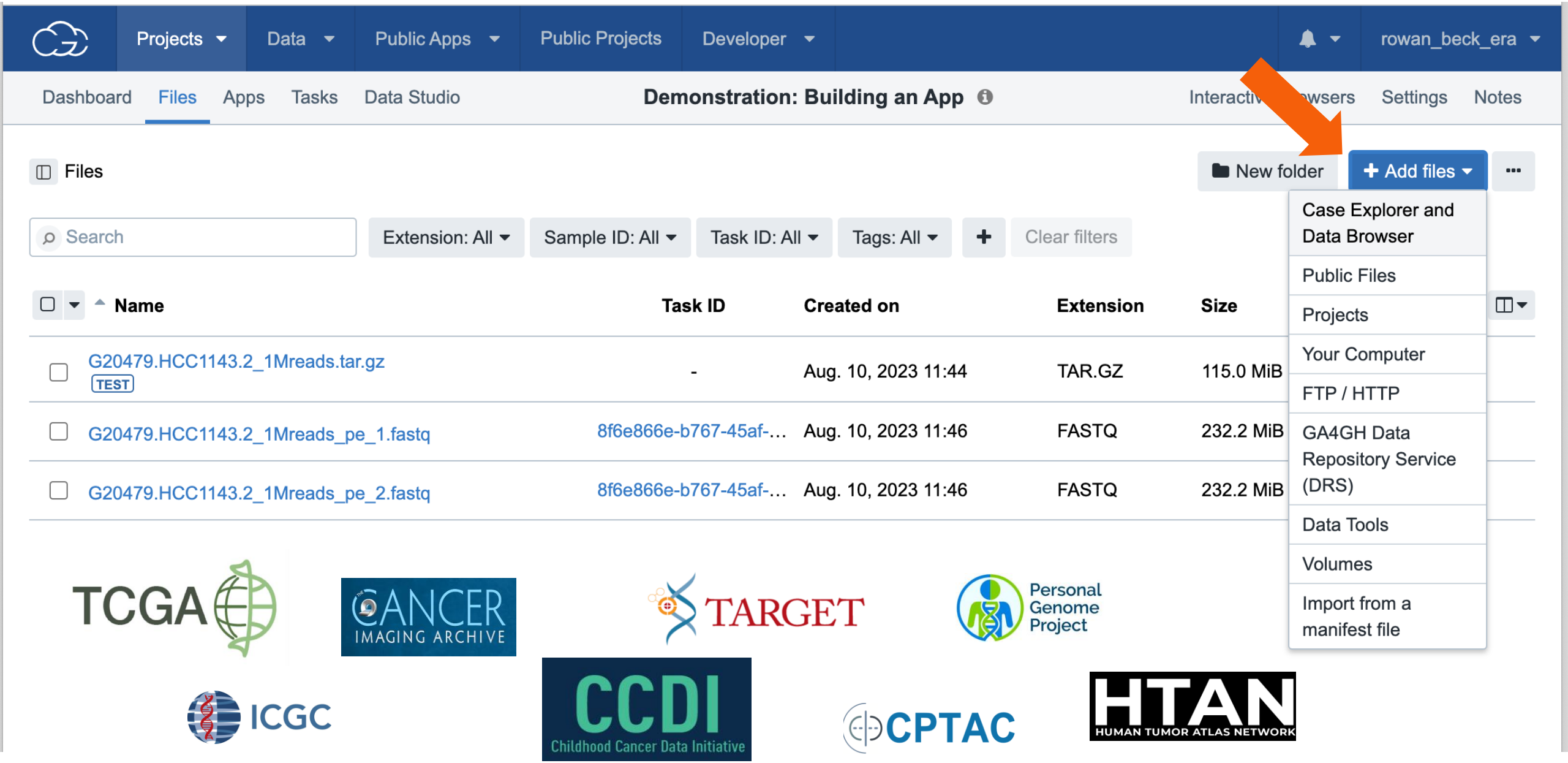
“Projects” are the main unit of Organization



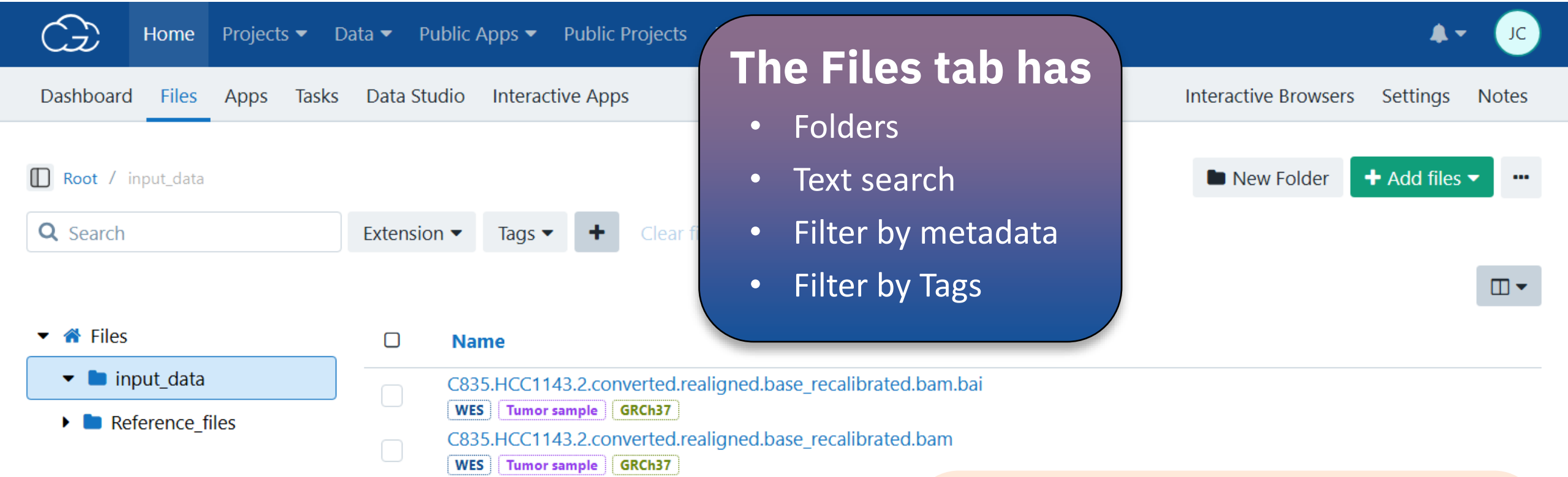
Projects can contain

- Files
- Apps and workflows
- Collaborators
- Interactive sessions
- Notes

Access Public Data, or Use Your Own



Files and file management



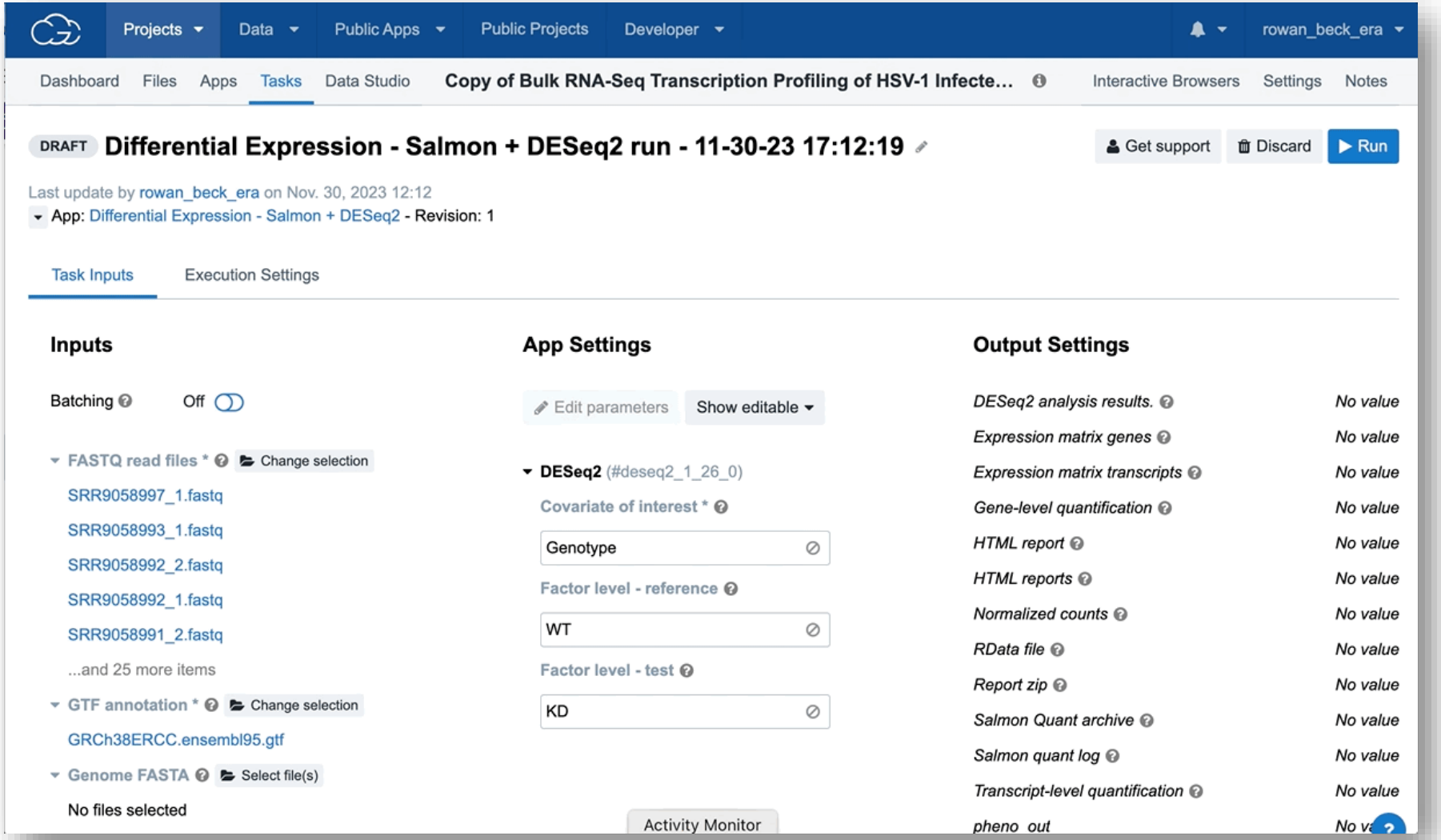
The Files tab has

- Folders
- Text search
- Filter by metadata
- Filter by Tags

LEARN

File Management:
https://youtu.be/Z4ygBEBunuY
Knowledge Center:
Files and metadata (cancergenomicscloud.org)

No coding required to run an analysis



Over 1300 curated tools and workflows

High quality apps & documentation

Customized user tools/workflows

Updated regularly

Optimized to run on cloud

Secure

Apps and Workflows: https://youtu.be/ayCKOafCGB0

Knowledge Center: Workflows and tools (cancergenomicscloud.org)

Data Studio – Interactive Data Analysis

Data Studio

Derive new insights using interactive analysis environments with Data Studio environments. Code in Python, SAS, and R to record and share your analyses.

Interact with files directly using our OHIF image viewer and Galaxy.

Videos:
R Studio on the CGC: https://youtu.be/_f8DWRTS0rE
JupyterLab on the CGC: https://youtu.be/iycWVHhqgXg

Documentation:
About Data Studio (cancergenomicscloud.org) | https://docs.cancergenomicscloud.org/docs/about-data-cruncher

VELSERA

Open Health Imaging Foundation

Apps – Bulk Processing Analysis

Command Tools and Workflows

- Convert any R or Python or bash script into an app that can be optimized and automated.
- Automated and optimize to scale up.

Videos:
CGC Onboarding:
https://youtube.com/playlist?
Public Apps Gallery:
https://youtu.be/ayCKOafCGB0?si=ps0O3ihGvgNvanB_

Documentation:
Apps on the CGC (cancergenomicscloud.org) | https://docs.cancergenomicscloud.org/docs/apps-on-the-cgc

CGC Democratizes Complex Analyses

A FAIR data ecosystem

- **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (FAIR)
- Immediate scaling (“Rent” vs “Buy”)
- Levels the playing field across institutions
- Many researchers can access data - no duplicate copies.
- Reproducibility - data and methods together.

HANDS-ON DEMO

cgc.SBGenomics.com

Variant Calling

VELSERA

Coming soon to the Cancer Genomics Cloud

Quantitative QC for Lung CT: 3D slicer

Dashboard | Files | Apps | Tasks | Data Studio | Interactive Apps | **Copy of COVID-19 Image Segmentation with Deep...** | Interactive Browsers | Notes

What it does

- Computes spatial overlap and boundary metrics
- Quantifies volume bias and over-segmentation
- Categorizes masks into pass, review, or fail
- Generates auditable segmentation quality records

Description

Email notifications

This project provides deep learning image seg
notebook, along with a pretrained model for s
images. The model was trained on a publicly a
Analysis files include:

- A Jupyter notebook walking through th
model. This notebook can be repurpos
applications by replacing the image pre
code and datasets.
- A Keras model based off of a UNet arch
encoder pretrained with ImageNet weig
- Images and masks in NIfTI (.nii) format,
masks for ground glass opacities and p
- Images and lung masks in NumPy array
images are shuffled from the order in th
- A plot of loss history over the course of t
will be overwritten if a new model is trained.
- Utility and dependency files.

To use the tools, select **Interactive Analysis > Data Cruncher > JupyterLab**. The notebook can be run using CPU, single-GPU, or multi-GPU instances in the **SB Machine Learning environment**. To use the model as is for predictions,

yachna
Copy, Write, Execute

yachnajain
Copy, Write, Execute

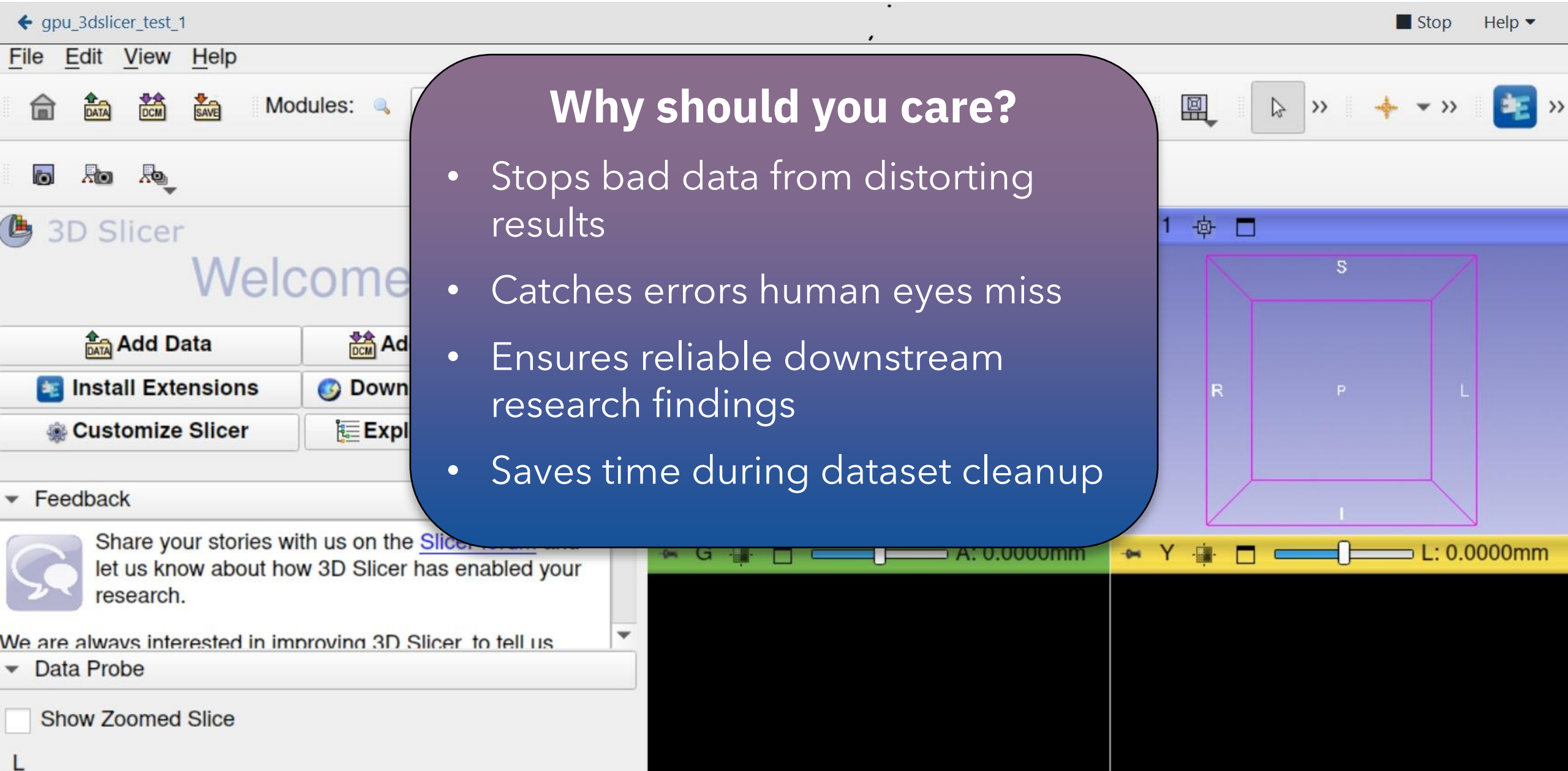
jenepher_cliffe_era_login
Copy, Write, Execute

Search

Tasks | Data Studio

COMPLETED **SBG Compressor CWL1.0 run - 06-29-26 07:48:48**
Submitted by: yatishp · Jun 29, 2026 3:48

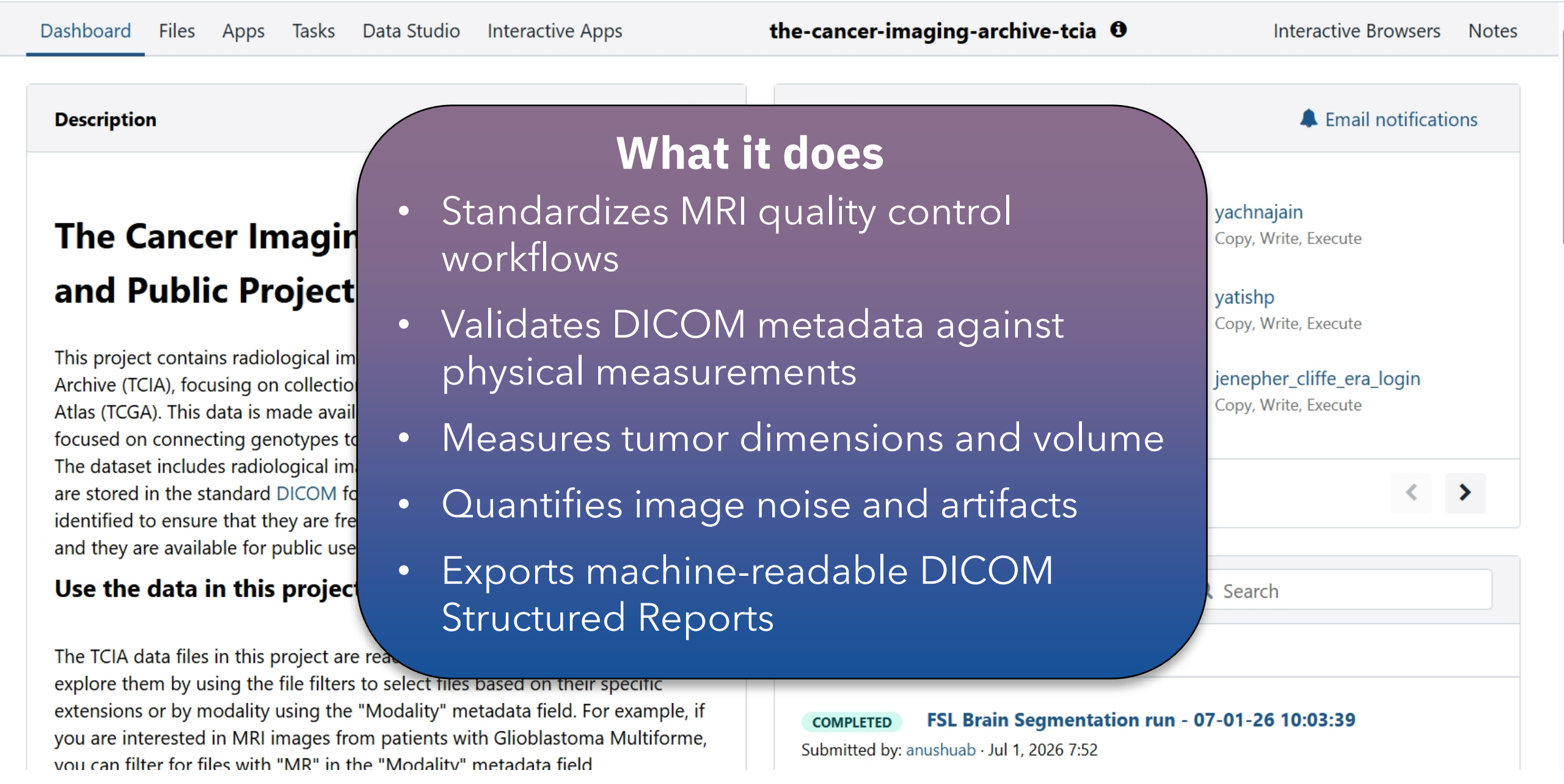
Quantitative QC for Lung CT: 3D slicer



Why should you care?

- Stops bad data from distorting results
- Catches errors human eyes miss
- Ensures reliable downstream research findings
- Saves time during dataset cleanup

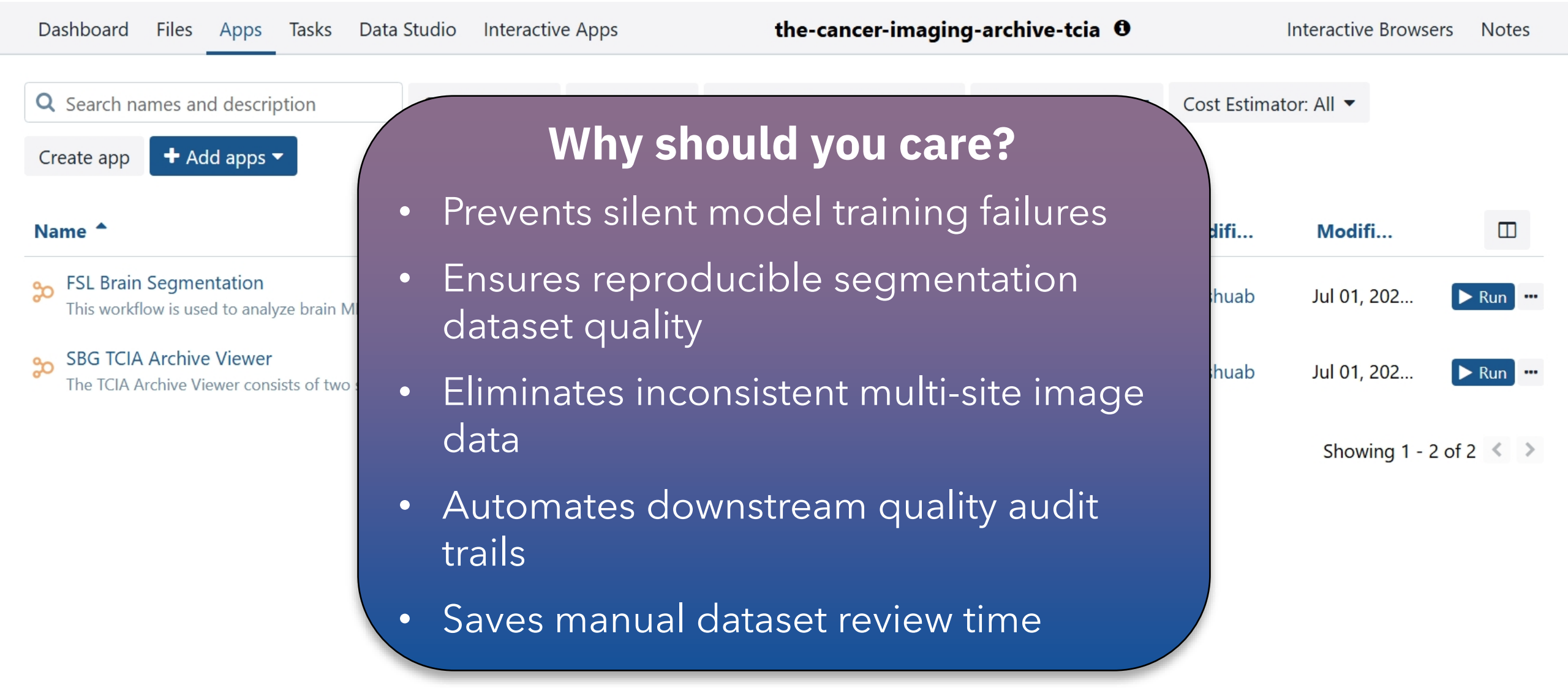
OHIF viewer: Data-Driven QC for Brain MRI



What it does

- Standardizes MRI quality control workflows
- Validates DICOM metadata against physical measurements
- Measures tumor dimensions and volume
- Quantifies image noise and artifacts
- Exports machine-readable DICOM Structured Reports

OHIF viewer: Data-Driven QC for Brain MRI



Why should you care?

- Prevents silent model training failures
- Ensures reproducible segmentation dataset quality
- Eliminates inconsistent multi-site image data
- Automates downstream quality audit trails
- Saves manual dataset review time

INTEGRATED TOOLS ENABLING

- Streamlined data access
- Reproducible
- High-throughput gating for AI pipelines
  (garbage in, garbage out protection)
- Democratized computational oncology

Resources for Working on the Cancer Genomics Cloud

You are not alone!

Onboarding resources

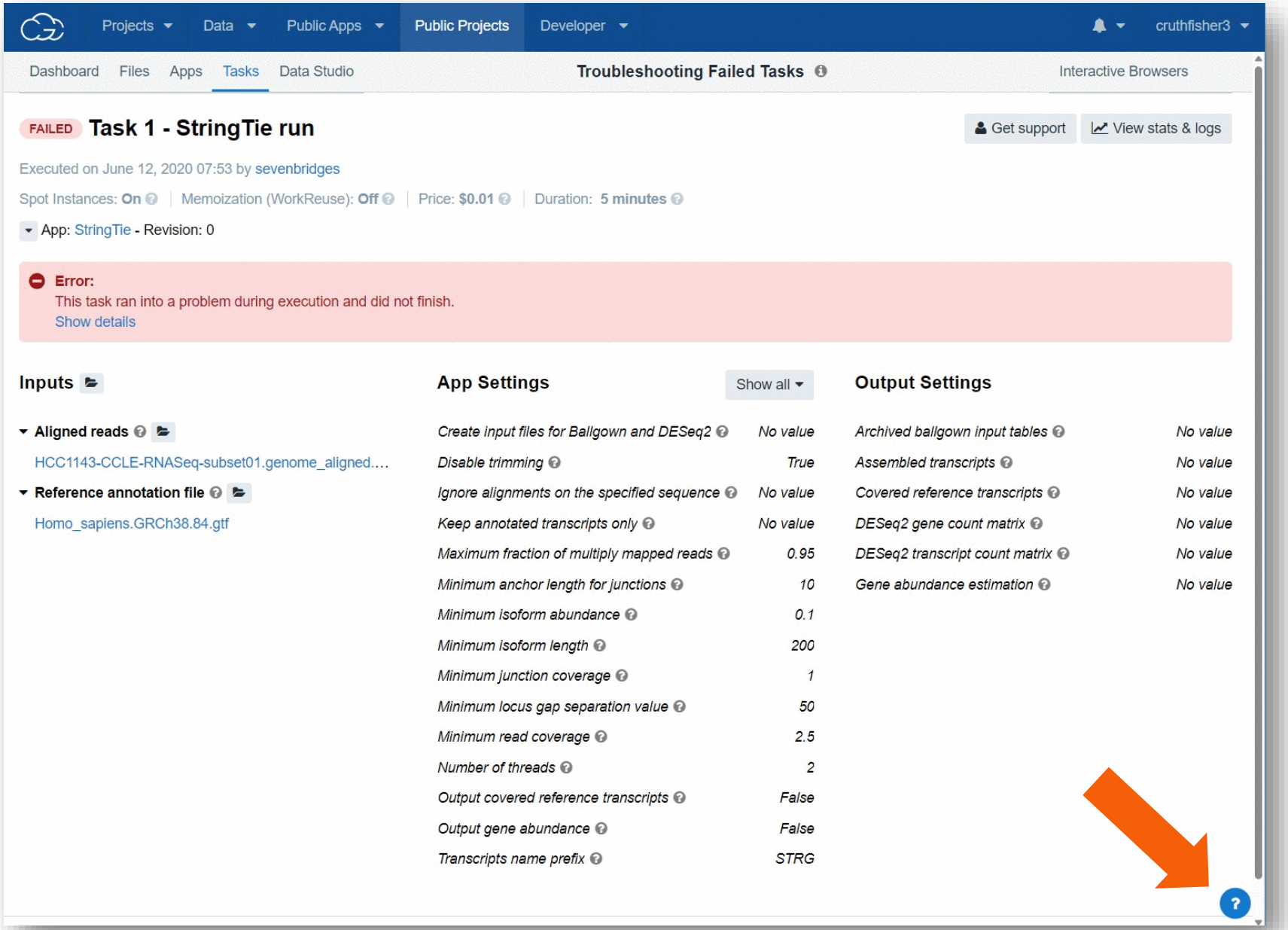
- **$300 pilot cloud credits (how to apply)**
- **Getting Started playlist on YouTube**
- **Extensive Knowledge Center**

Public apps and projects

- Curated tutorials and templates to adapt to your data.
- Vetted and tested apps and tools

New Public Projects (Template projects)	37
Public Apps (Published via projects)	33
Public Shiny-based Apps (PUBLIC_WEB_APP)	15

Get The Support You Need



Get The Support You Need

Every Week:

- 10:00 am ET Tuesday
- 2:00 pm ET Thursday

NCI Funding Is Available on the CGC

Pilot Credit Funds

- **$300 of cloud credits**
- Free for new CGC users
- Easy to request when signing up (see Getting Started – Cancer Genomics Cloud)
- Fast approval

NIH NATIONAL CANCER INSTITUTE

Collaborative Projects

- Cost estimation, optimization, and planning support
- Great for researchers new to bioinformatics and cloud approaches
- **Up to $10k compute/storage costs**
- Fast, rolling applications
- To date > 60 projects
- Apply at Collaborative Projects – Cancer Genomics Cloud

Stay in touch.

Support@Velsera.Com

https://meet.google.com/kbs-ojnj-dcg

Every Week:

- 10:00 am ET Tuesday
- 2:00 pm ET Thursday

Jeni Cliffe
Community Engagement Manager
Jenepher.Cliffe@velsera.com

Michelle Engle, PhD
Community Engagement Manager
Michelle.Engle@velsera.com

Thank you.

Learn more at
CancerGenomicsCloud.org

VELSERA

COMING SOON!
cgc.SBGenomics.com

OHIF viewer and 3D slicer

VELSERA