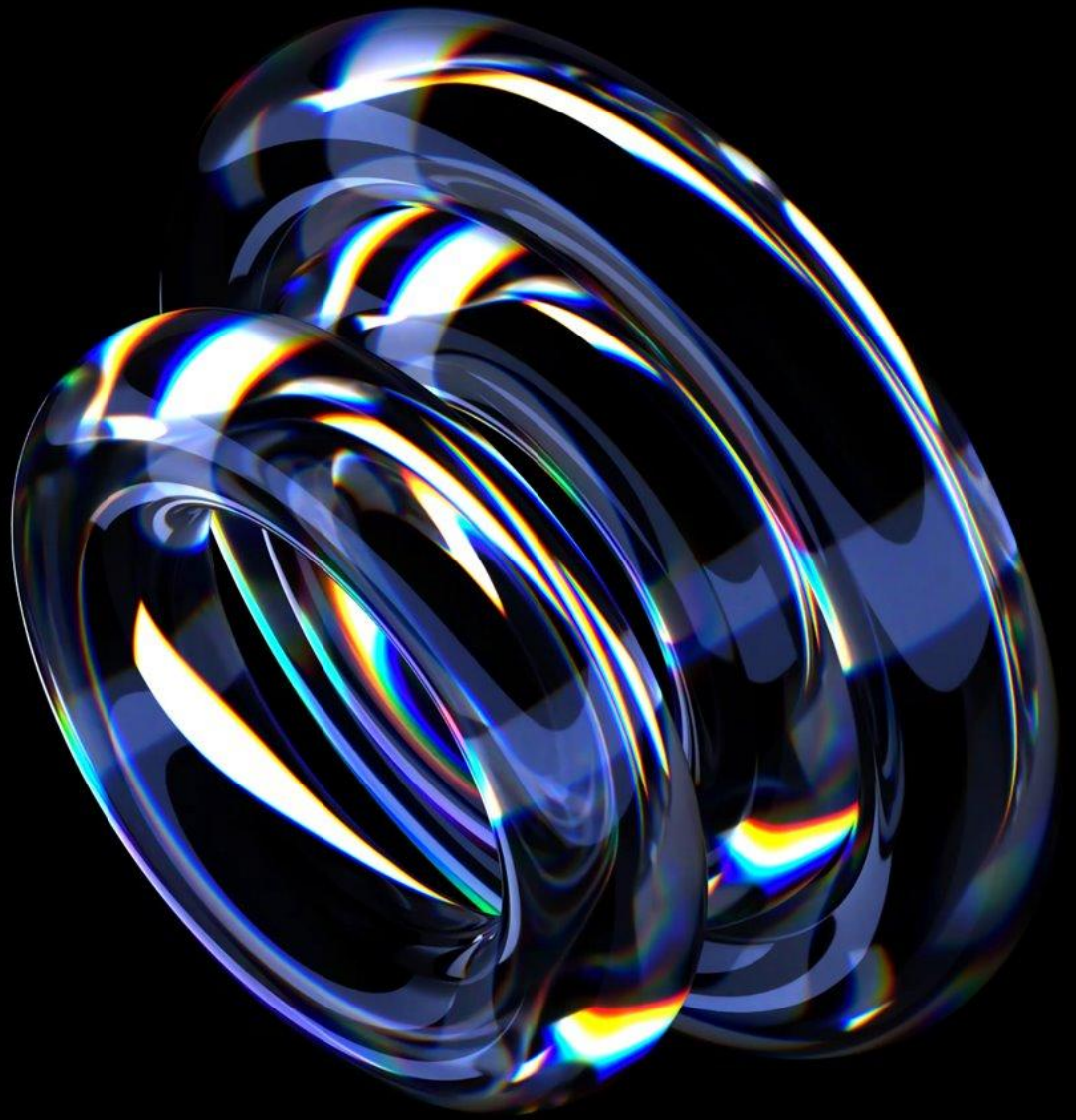
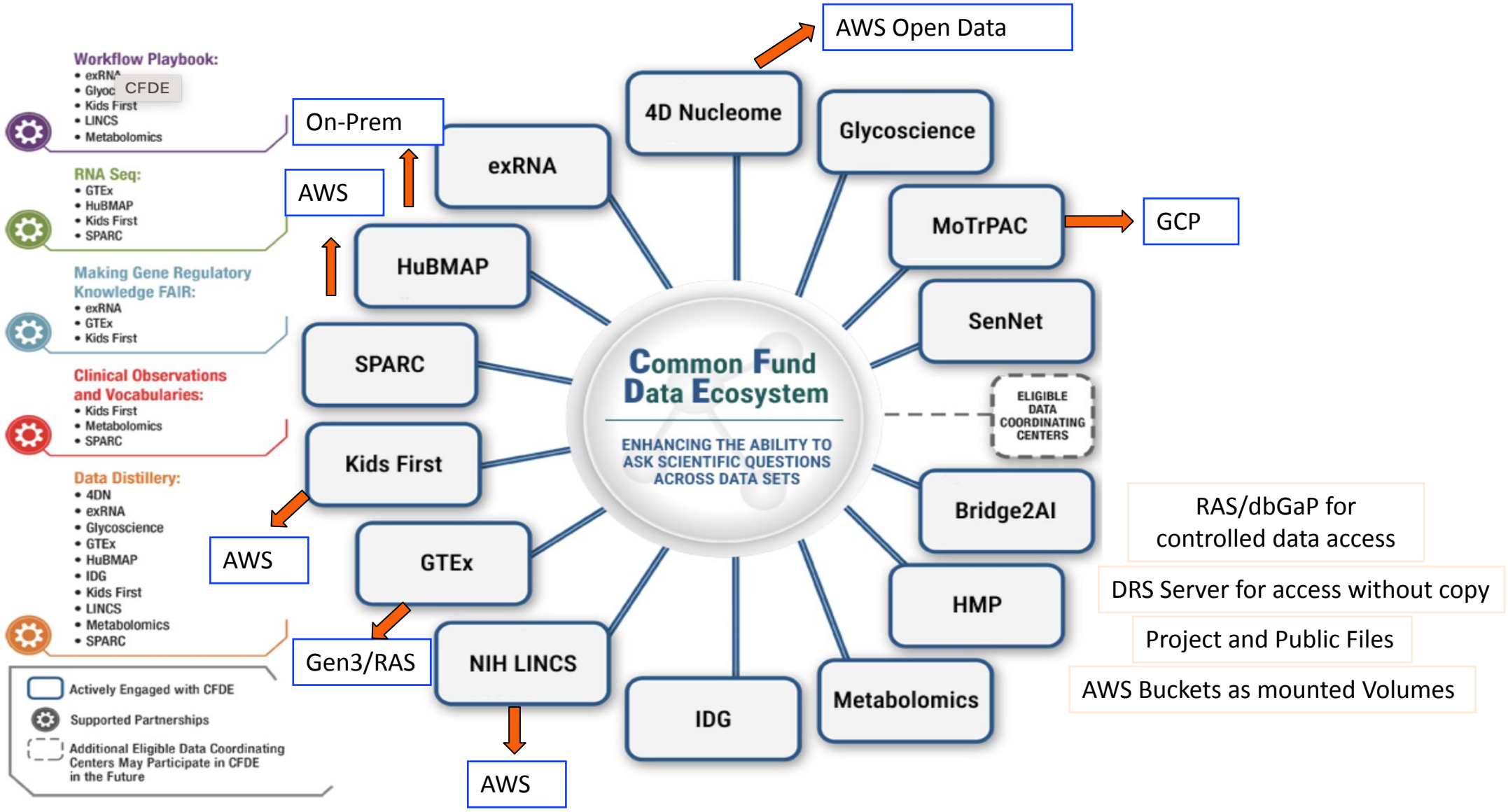
Cloud Workspace Pilot.

CFDE All Hands Training

Sangeeta Shukla
Eric Tobin
20 March 2024



VELSERA


Workflow Playbook:
• exRNA
• Glyoc
• Kids First
• LINCS
• Metabolomics
CFDE
RNA Seq:
• GTEx
• HuBMAP
• Kids First
• SPARC
Making Gene Regulatory Knowledge FAIR:
• exRNA
• GTEx
• Kids First
Clinical Observations and Vocabularies:
• Kids First
• Metabolomics
• SPARC
Data Distillery:
• 4DN
• exRNA
• Glycoscience
• GTEx
• HuBMAP
• IDG
• Kids First
• LINCS
• Metabolomics
• SPARC
Actively Engaged with CFDE
Supported Partnerships
Additional Eligible Data Coordinating Centers May Participate in CFDE in the Future
On-Prem
AWS
AWS
Gen3/RAS
AWS
AWS Open Data
GCP
RAS/dbGaP for controlled data access
DRS Server for access without copy
Project and Public Files
AWS Buckets as mounted Volumes
exRNA
4D Nucleome
Glycoscience
MoTrPAC
SenNet
HuBMAP
SPARC
Kids First
GTEx
NIH LINCS
IDG
Metabolomics
HMP
Bridge2AI
ELIGIBLE DATA COORDINATING CENTERS
Common Fund Data Ecosystem
ENHANCING THE ABILITY TO ASK SCIENTIFIC QUESTIONS ACROSS DATA SETS

Big Data Sourcing and Management within CAVATICA

- Address per project storage limitations
- Remove redundancy in ELT performance
- Projects can handle files found in different storage space
- Easier retrieval of Public Files
- Enable re-use by multiple projects/pipelines running simultaneously
- Controlled access at multiple points
- Processed result files can be reused
- Data Studio to allow data manipulation

CAVATICA

CAVATICA is a **data analysis** and **sharing platform** designed to accelerate **discovery** in a **scalable**, **cloud-based** compute environment where **data**, **results**, and **workflows** are **shared** among the world's research community. Developed by Seven Bridges and funded in-part by a grant from the National Institutes of Health (NIH) Common Fund, CAVATICA is continuously updated with new **tools** and **datasets**.

Our goal is to help researchers **collaborate**, **share**, **interoperate**, and **connect** with any and all other data ecosystems in order to empower data analysis across diseases, ages and geography. By connecting previously disconnected datasets, CAVATICA supports researchers and patients across the United States and throughout the world to generate new insights into pediatric diseases.

The Cavatica Platform has been funded in whole or in part with Federal funds from the National Institutes of Health, Contract No. U2C HL138346 and funding from the Center for Data-Driven Discovery in Biomedicine (D^3b)

CAVATICA

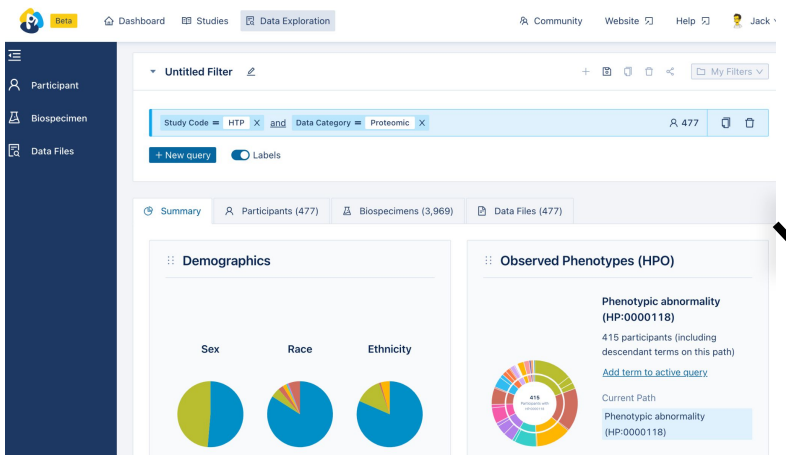
CAVATICA is a **data analysis** and **sharing platform** designed to accelerate **discovery** in a **scalable**, **cloud-based** compute environment where **data**, **results**, and **workflows** are **shared** among the world's research community. Developed by Seven Bridges and funded in-part by a grant from the National Institutes of Health (NIH) Common Fund, CAVATICA is continuously updated with new **tools** and **datasets**.

Our goal is to help researchers **collaborate**, **share**, **interoperate**, and **connect** with any and all other data ecosystems in order to empower data analysis across diseases, ages and geography. By connecting previously disconnected datasets, CAVATICA supports researchers and patients across the United States and throughout the world to generate new insights into pediatric diseases.

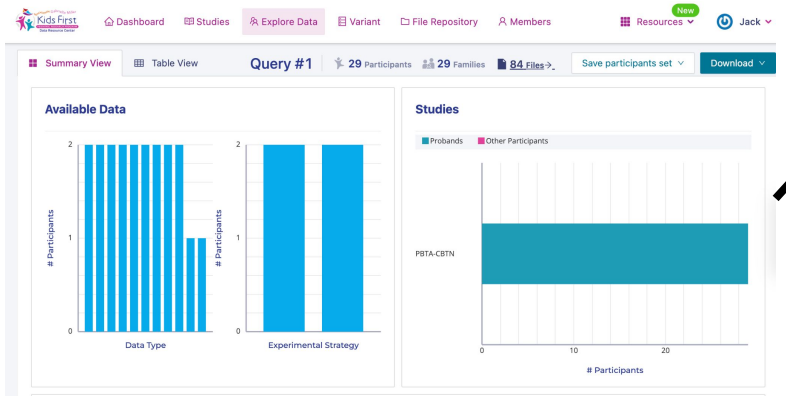
The Cavatica Platform has been funded in whole or in part with Federal funds from the National Institutes of Health, Contract No. U2C HL138346 and funding from the Center for Data-Driven Discovery in Biomedicine (D^3b)

Empowering a one-to-many integration in the cloud

INCLUDE



Kids First



CAVATICA

TOPMed

AnVIL

TCGA

TARGET

PDC

ICDC

UDN

NIH National Heart, Lung, and Blood Institute | BioData CATALYST Powered by Seven Bridges

AnVIL

CANCER GENOMICS CLOUD SEVEN BRIDGES

CANCER GENOMICS CLOUD SEVEN BRIDGES

CANCER GENOMICS CLOUD SEVEN BRIDGES

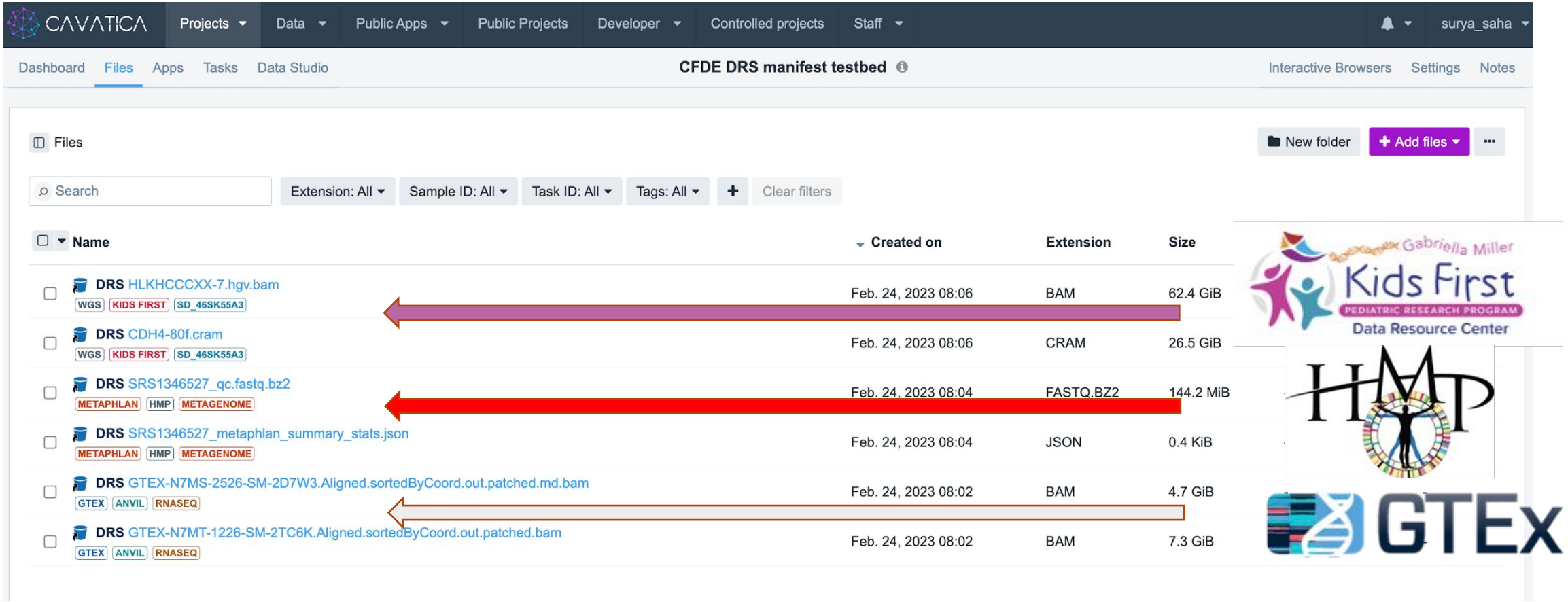
CANCER GENOMICS CLOUD SEVEN BRIDGES

db GaP genotypes and phenotypes

Interoperability stack has provided the ability to have datasets from different "places" together in single platform

VELSERA

Interoperable data on CAVATICA via DRS



- The DRS client on CAVATICA resolves authorizations
- Researchers can focus on the science and use them as input to a task

Office Hours

Collaborative Project Interest Form

Q&A and … Thank You!

Sangeeta Shukla,
Kids First / Children's Hospital of Philadelphia

Eric Tobin,
Velsera / Seven Bridges CAVATICA

Appendix: Other References

Who are the CAVATICA Users?

The platform 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 Push-button Automation Solutions**
- **Analyze Data at Scale with Tools and Workflows**
- **Conduct Interactive Exploratory Analyses**
- **Explore/Visualize Results and Gather Insights**
- **Easily Collaborate with Other 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 with Upstream/ Downstream Systems**