Breakout Session 1: Track A

# Using Seven Bridges' CAVATICA to Empower Use of the INCLUDE DCC Platform

Dr. Jack DiGiovanna
*Chief Science Officer, Velsera*

INCLUDE
Data Coordinating Center

# *Using Seven Bridges' CAVATICA to Empower Use of the INCLUDE DCC Platform*

Jack DiGiovanna, PhD

January 17th, 2024

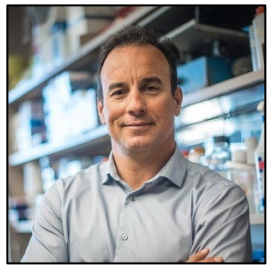
# Our Mission

Our mission is to accelerate discoveries that continuously improve and enrich the lives of people with Down syndrome through accessible data resources that enable interdisciplinary collaboration, connected communities, and the expansion of a global Down syndrome knowledge ecosystem.

more info: https://includedcc.org/

# Data Hub combines expertise to advance research

**AOC**



**Espinosa**

LINDA CRNIC INSTITUTE for DOWN SYNDROME

**DMC**

**O'Connor**

SageBionetworks

Haendel

University of Colorado Anschutz Medical Campus

Carroll

VANDERBILT UNIVERSITY MEDICAL CENTER

**DPC**

**Resnick**

Children's Hospital of Philadelphia
Center for Data Driven Discovery in Biomedicine

Ferretti

CHU Sainte-Justine
Mother and Child University Hospital Centre
Université de Montréal

DiGiovanna

SevenBridges

VELSERA

**NIH Project Scientist: Dr. Huiqing Li, NHLBI**

Charlene Schramm, NHLBI
Lanre Ojese, NHLBI
Melissa Parisi, NICHD
Sujata Bardhan, NICHD
Rachel Dows, NICHD
Laurie Ryan, NIA
Gail Pearson, NHLBI
Valerie Cotton, NICHD
Linda Garcia, NICHD
Erika Tarver, NIA
Sarah Ashraf, NICHD
Erica Landis, NIH/OD

A cooperative agreement with NIH

NIH National Heart, Lung, and Blood Institute

NIH *Eunice Kennedy Shriver* National Institute of Child Health and Human Development

NIH National Institute on Aging

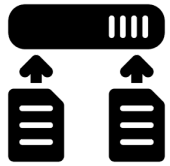
And many others…

# INCLUDE Data Hub is guided by our mandate

Should ***ACCELERATE*** the discovery process via **platform-based approaches** that leverage the cloud and other data-driven technologies across a diverse user community.

Drive **collaborative discovery** and de-siloed research and clinical translation across the DS condition co-occurrence landscape.

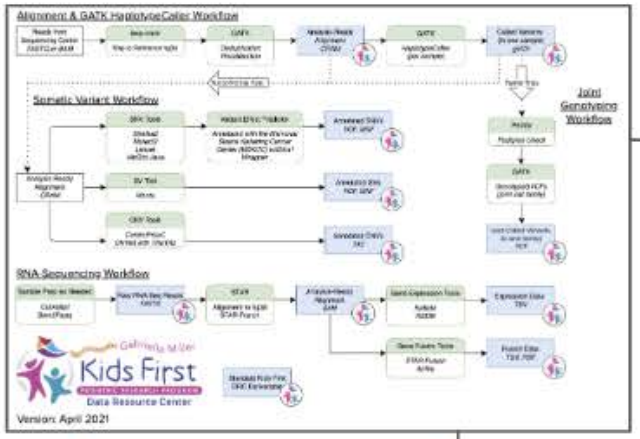
Develop and implement leading edge technologies that enable NIH-based frameworks of data **security and governance** while providing a seamless experience of end users. Support the integration of multiple **types** of data **sources**.



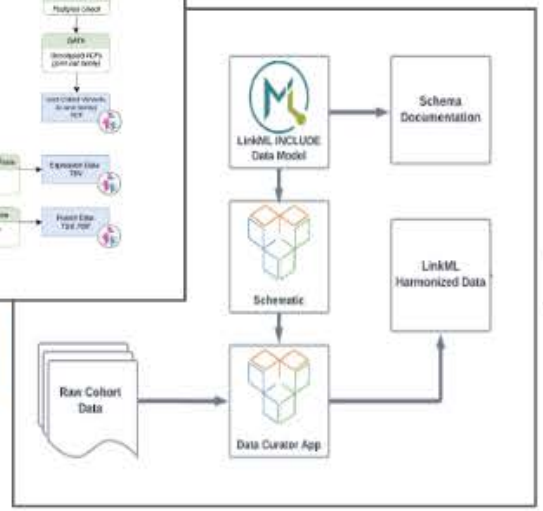
Support the integration of multiple **types, modalities,** and **sources** of data spanning a diverse research landscape.

# INCLUDE Data Hub Portal – Data-Driven Discovery

**Source Data** | **Harmonized Data** | **Analysis Results**




SYNAPSE




HL7 FHIR

Clinical variable metadata

aws

Source Files
Primary Harmonized Files
Secondary Matrices

INCLUDE Data Hub

CAVATICA

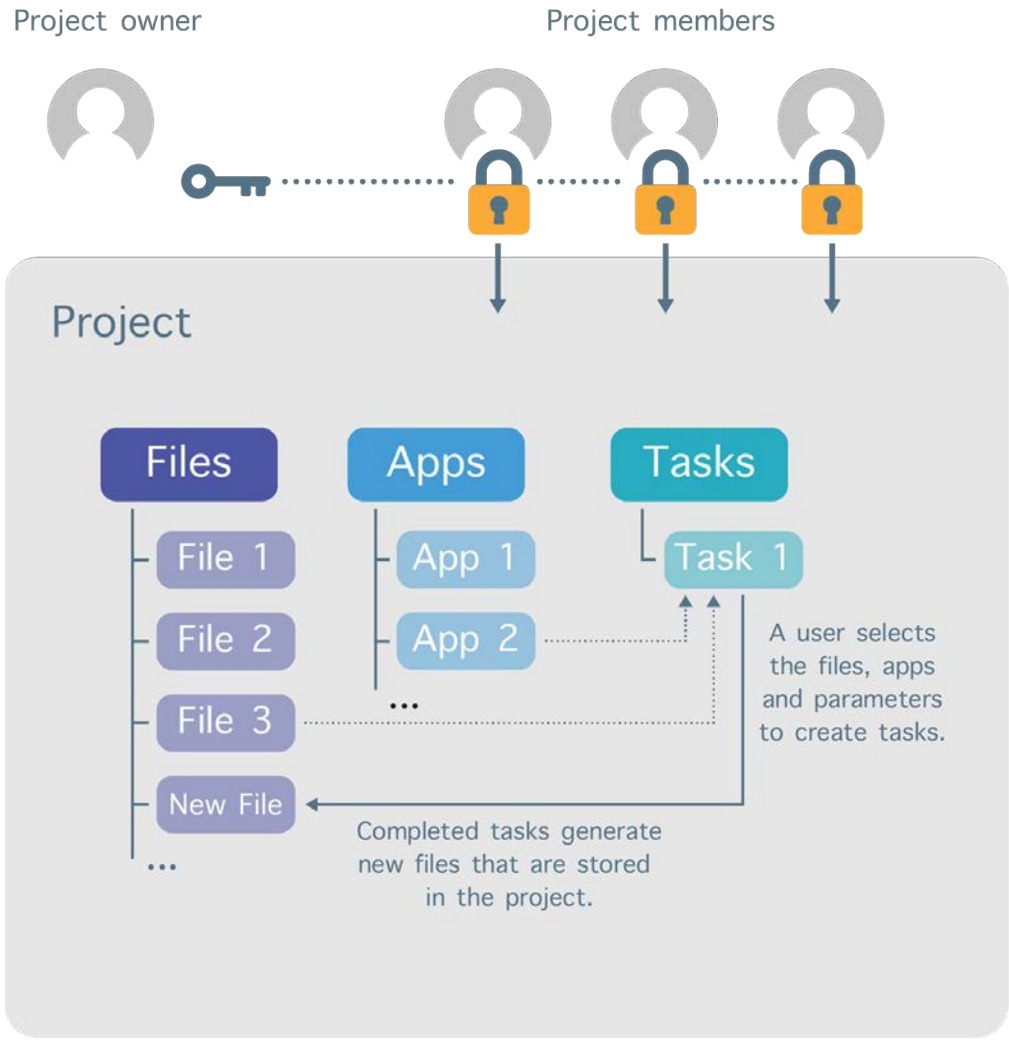
# CAVATICA

# Collaborate and analyze INCLUDE data in a workspace

Projects help researchers collaboratively analyze data

Structure and permissions enable interoperability

Costs managed & controlled within projects; integrated with STRIDES reporting



# Support for a spectrum of analysis needs & personas

## Production pipelines & scaling

Mature analysis pipelines are optimized, validated, & auditable. These pipelines can leverage metadata and intelligent scheduling.

COMMON WORKFLOW LANGUAGE

docker

nextflow

## Prototyping & Interactive Analysis

Iterative experimentation can be key to discovery. Visualization & data science happen here.

RStudio®

jupyter

## Custom Webapps

Ultimately results often take the form of a custom figure, table, or report that is quickly digestible and shareable.

Shiny

Dash

# INCLUDE Cloud Credit Program

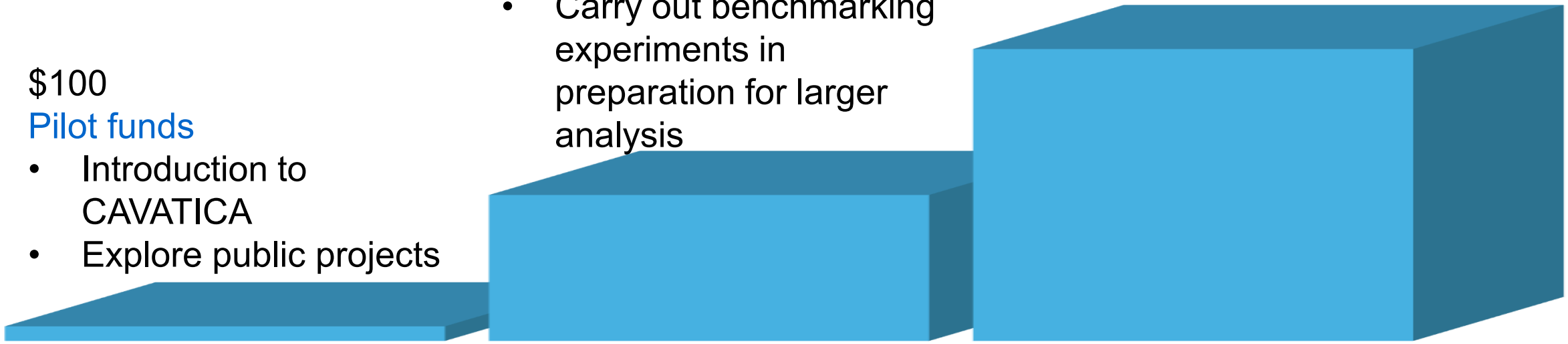
# INCLUDE Cloud Credit Program

$100
Pilot funds
- Introduction to CAVATICA
- Explore public projects

$1,000
Limited
- Build and test workflows
- Complete small analysis projects
- Carry out benchmarking experiments in preparation for larger analysis

Up to $20,000
Extended
- Larger research projects

*Applications must include preliminary results and benchmarking outcomes from the Limited Credits



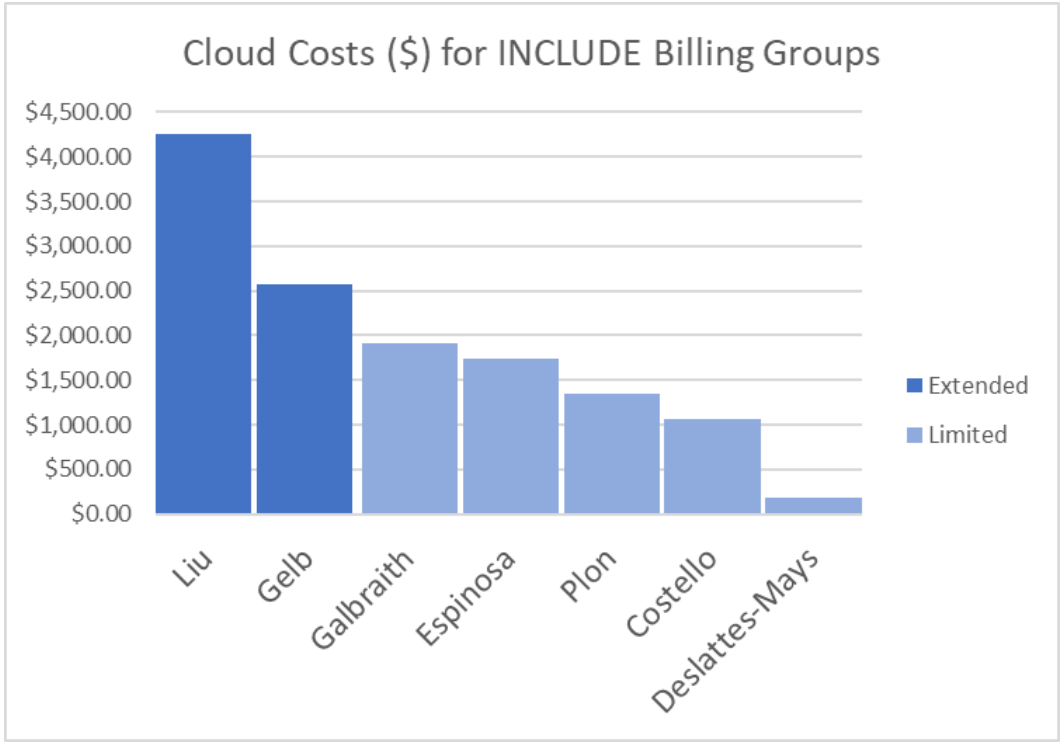
# INCLUDE users are starting to leverage CAVATICA

- 308 INCLUDE users supported by INCLUDE Cloud Credits
  - 50 users with eRA Commons
    - Accessed Cavatica an average of 14 times
    - Created 135 user projects and are a part of 159 user projects

Supported DS community

- Billing Group configured to meet needs, track usage, and help control cloud costs

- 1 Hackathon (**BioIT**)
- 4 Trainings (**DS3**, **Elements of Style**) twice
- Trained 20 students per year in DS3 workshop
- ~40 registered for Elements of Style and attended at least one day



# Leverage the Data Hub for teaching

NIH strategic priority to train the next generation of data scientists

INCLUDE and DCC investments **have made multiple datasets "research ready"**

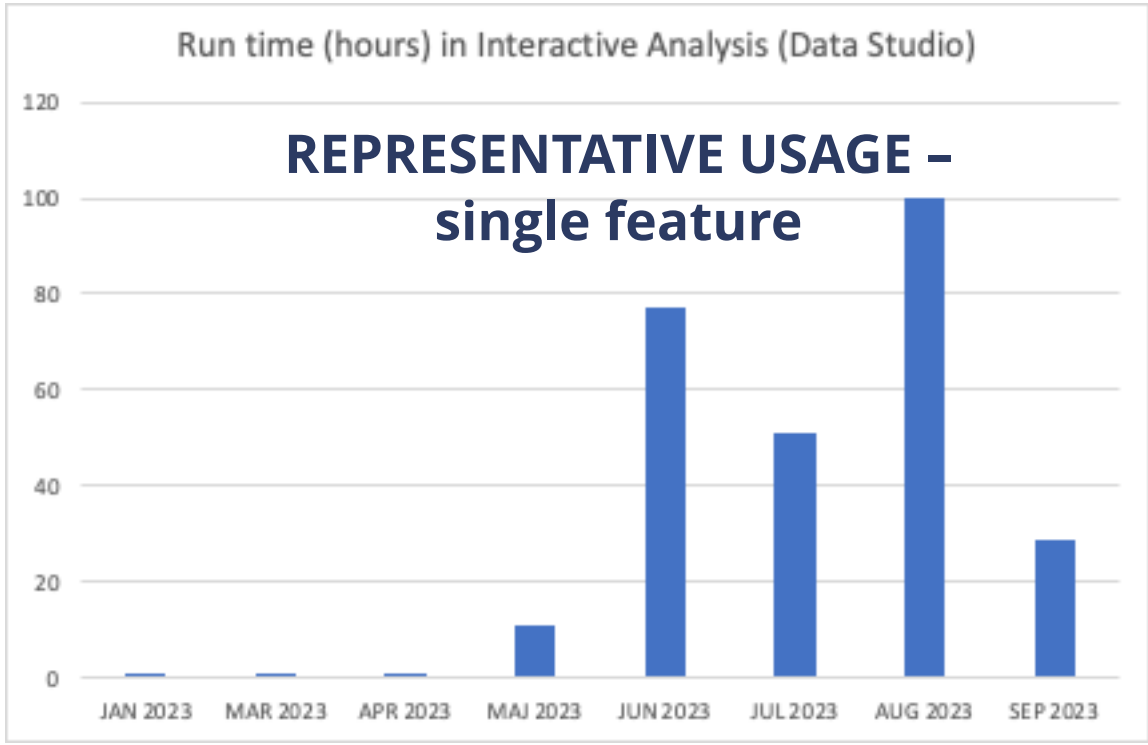
- **D**ata **S**cience for **D**iverse **S**cholars in **D**own **S**yndrome Research (DS3) workshop

Support, tools, collaboration, admin built in

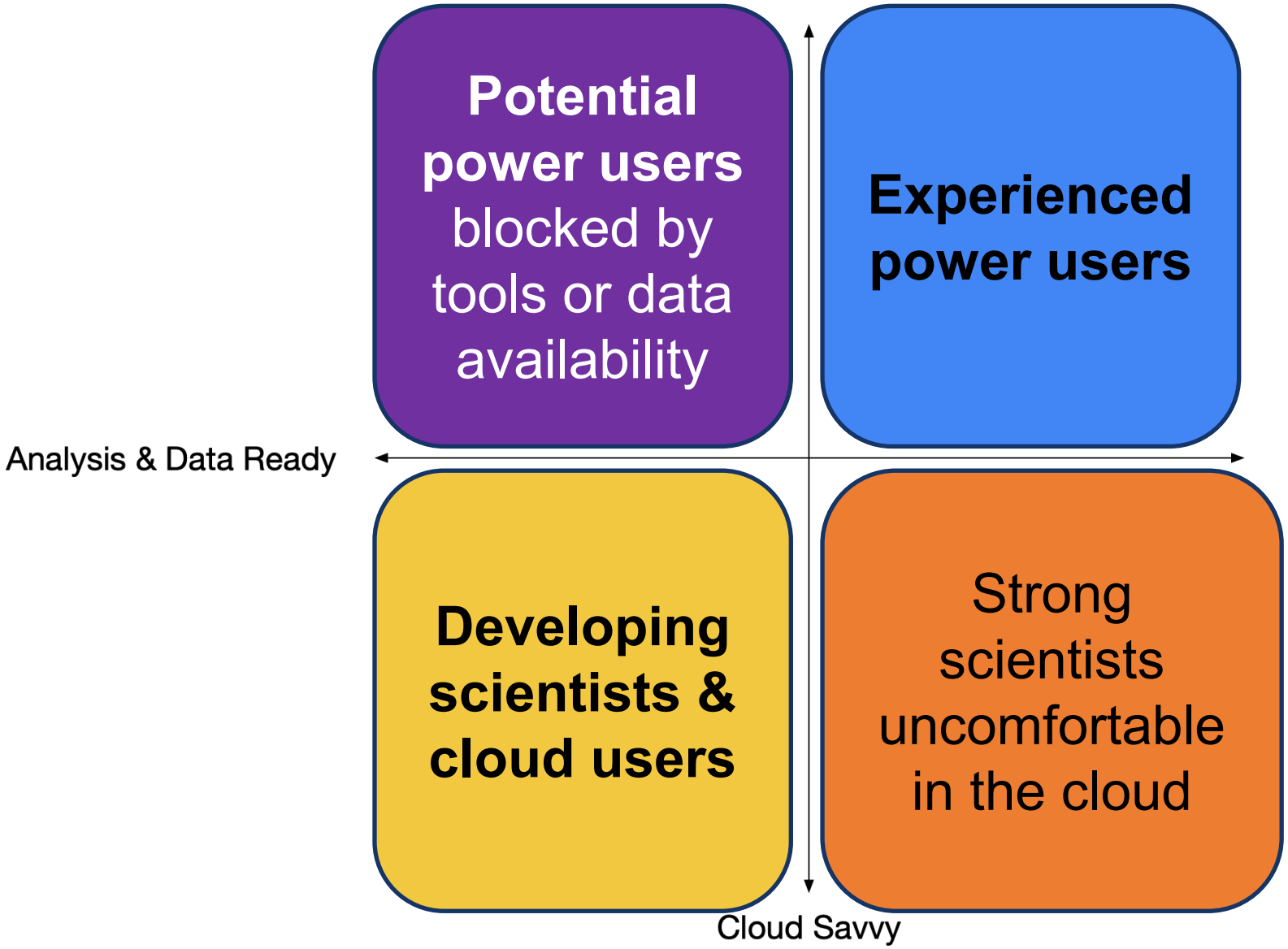
INCLUDE Cloud Credits helped researchers get started

# Lessons from usage of Cloud Credits

- People **take time** to use their credits
- Users **contact us late** in the research analysis **life cycle**
- Research (& data) **comes in waves**
- Predefined user personas and support strategies **can be limiting**
- **Adoption of new technology** while trying to **complete research** is **challenging**



# Leverage STRIDES to accelerate diverse users



# Supporting current and future awardees

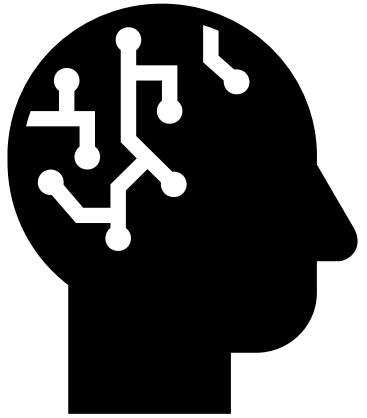
Work with the Data Hub to **encourage more users** to accelerate their research via STRIDES generous funding

Help current awardees to **maximize outcomes**

**Tailor support** to different user clusters

Refine user metrics

- Datasets have reached a state of maturity to facilitate larger collaborations
- We will on-board research labs while their data is being processed
- Set up regular touchpoints to check on progress of user groups
- Encourage peer-mentoring and sharing of best practices



Thanks to **Jared Rozowsky**, **Surya Saha,** & **David Higgins** for driving this work

jared.rozowsky@velsera.com

# Open Discussion Q&A