

# Welcome to the Common Fund Data Ecosystem Webinar

Covering:

Common Fund Data Ecosystem (CFDE) Project  
Science and Technology Research Infrastructure for Discovery,  
Experimentation, and Sustainability Initiative (STRIDES)



National Institutes of Health  
*Office of Strategic Coordination - The Common Fund*

# Agenda



*The Common Fund*

- Welcome and introductions
- Common Fund goals for data resources
- Common Fund Data Ecosystem Project (CFDE)
- Science and Technology Research Infrastructure for Discovery Experimentation and Sustainability (STRIDES) Initiative
- What these projects mean for you/ your program
- Q&A
- Where to go with questions

# After this webinar you should...



*The Common Fund*

- Understand opportunities that CFDE and STRIDES aim to address
- Understand Common Fund's goals when it comes to CFDE and STRIDES
- Understand your role in both projects moving forward
- Know how to leverage STRIDES
- Know who may contact you about the CFDE
- Know who you can contact with questions

# Welcome and Introductions!



**National Institutes of Health**  
*Office of Strategic Coordination - The Common Fund*

# Panelists



*The Common Fund*

Betsy Wilder, Director,  
Office of Strategic  
Coordination

Jennifer Yttri, Project Manager,  
Common Fund Data Ecosystem,  
MITRE

Simon Twigger, Senior  
Scientific Consultant,  
BioTeam



# Common Fund's Vision for Data Resources



*The Common Fund*

Common Fund Programs are intended to provide resources that accelerate discovery across many different biomedical research fields.

Often these resources include large datasets and the digital tools needed to mine and analyze data.

To maximize impact, these datasets must be leveraged by researchers across many different disciplines, with varying expertise in bioinformatics and large-scale data analysis.

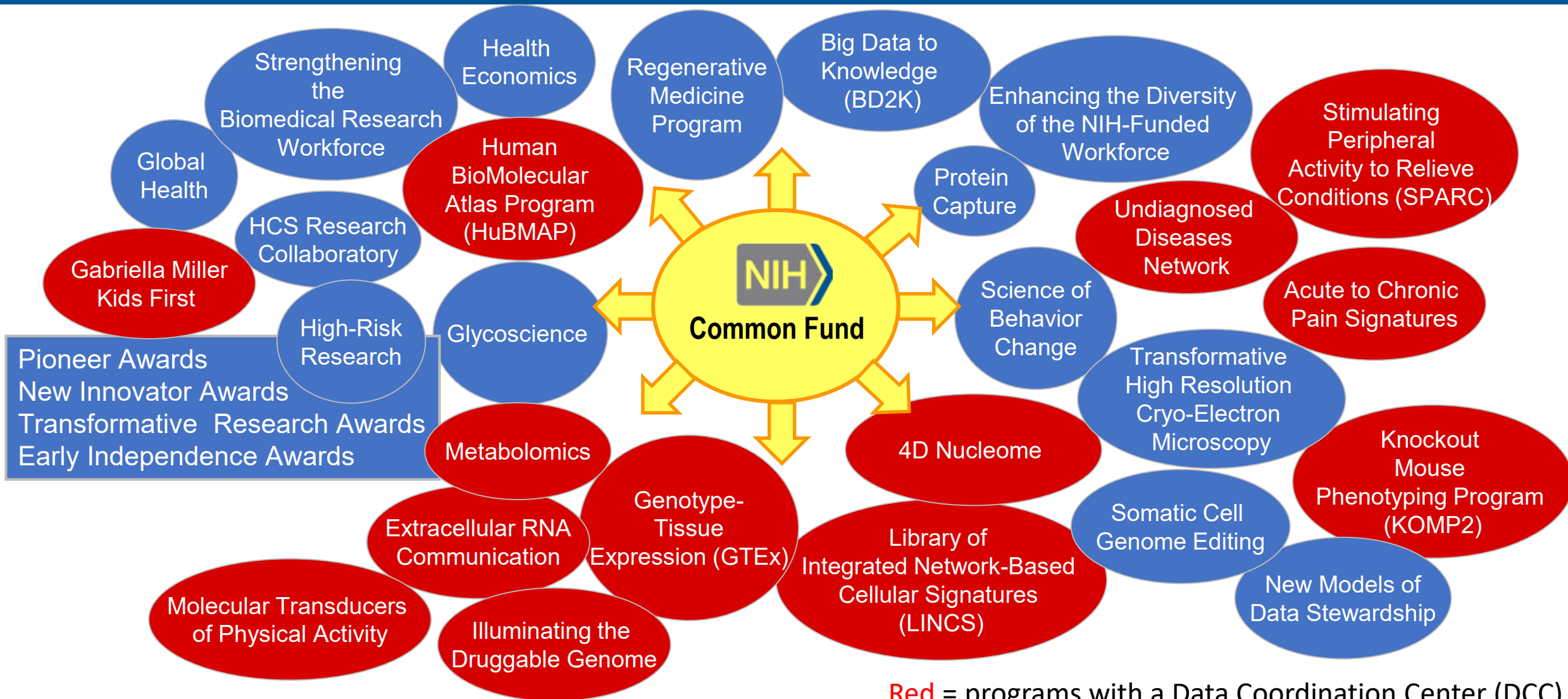
In addition, these datasets must be usable together across interoperable platforms.



# Common Fund Programs – FY19



The Common Fund



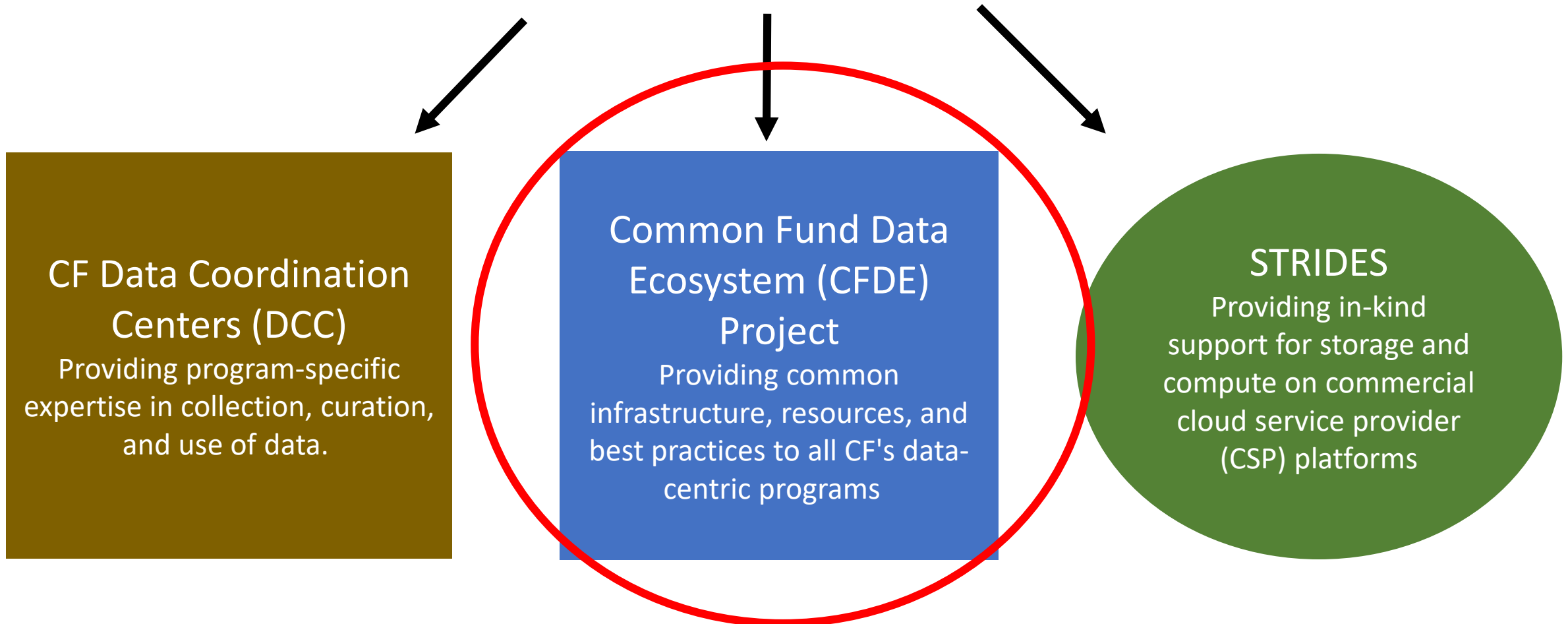
Red = programs with a Data Coordination Center (DCC)

# How is the Common Fund Addressing Challenges?



The Common Fund

## Common Fund Support





# The Common Fund Data Ecosystem (CFDE) Project



National Institutes of Health  
*Office of Strategic Coordination - The Common Fund*

# Common Fund Data Ecosystem: Goals



*The Common Fund*

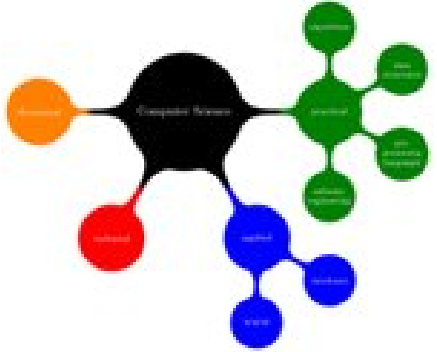
The goals of the CFDE are to advance science by:

- Enhancing the ability to ask scientific questions across data sets.
- Enabling the uptake, reuse, and addition of Common Fund data and tools from future, current, and ended programs.
- Supporting the storage, sharing, and sustainability of Common Fund Data Sets.
- Providing training that maximizes scientists' ability to upload data and use Common Fund data and other resources.

# To get there, CFDE will work on:

- ✓ Providing a catalog of CF data sets with a gateway to access them.  
X It won't create a platform to host the data.
- ✓ Facilitating adoption of community-developed standards and APIs/tools for queries across multiple data sets.  
X It won't create a full stack to access and analyze data.
- ✓ Facilitating data usage within a cloud environment through technical support and training.
- ✓ Providing basic services to start up new DCCs and to sustain data sets and tools stored in the cloud after a CF program ends.  
X It won't support ongoing data collection.
- ✓ Working with DCCs to pilot and accomplish these activities. DCCs will have the opportunity to apply to be part of the CFDE in FY20.





## What we know...

- In FY2020, CF awarded Owen White an OTA to coordinate CFDE activities.
- The CFDE CC works as two teams:
  - Engagement Team—builds partnerships with DCCs, communicates to user communities, and supports training.
    - Titus Brown, RTI
  - Technical Team—provides technical expertise; develops tools and standards to be implemented by DCCs, including data/metadata management, authentication/authorization, and FAIR.
    - Avi Ma'ayan, Carl Kesselman, Ian Foster, Brian O'Connor
- Scientific opportunity! The goal is to make Common Fund datasets more useful alone and in combination. The CFDE CC is working with CF DCCs to learn from their experiences and provide technical support for common challenges.



## What we don't know...

- Specifics of exercises with DCCs—this will depend on what DCCs want to do.



## What we know...

- DCC Funding! In FY2020, the Common Fund plans to announce an opportunity for its DCCs to participate in the CFDE.
  - Awards are estimated at \$250K/yr/award, with up to 10 awards in FY2020.
- Participation will include engagement with other CF DCCs and the CFDE.
- Leveraging STRIDES is important, and there's a programmatic role, but it is separate from CFDE.
  - Several programs are working with STRIDES and have lessons learned to share.

## What we don't know about the new DCC funding opportunity

- Exact timing of the CFDE awards, other than FY2020.
- Award mechanism (whether supplement or something else).
- How many awards will be made (up to 10).
- What DCCs will propose to do (i.e., need to look out for overlap).

# Status of CFDE Project activities



*The Common Fund*

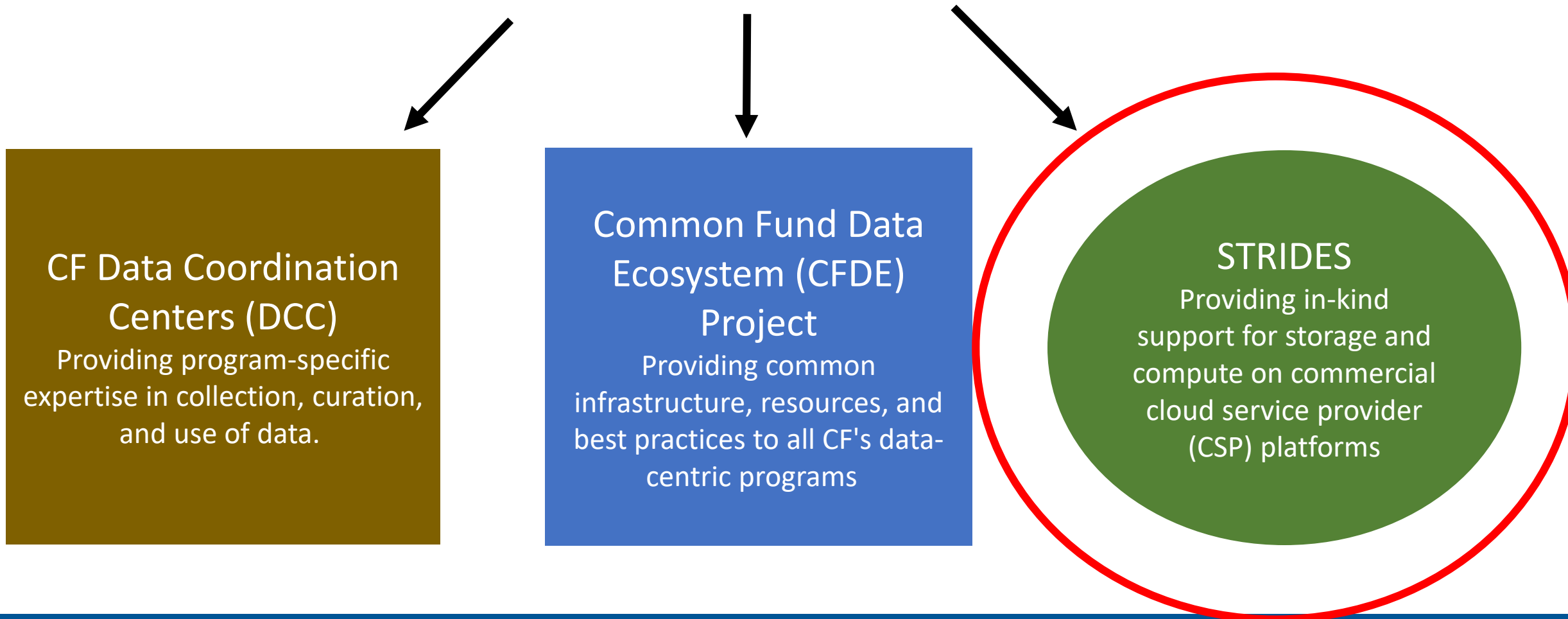
- **Funded the CFDE Coordinating Center (Owen White, PI)**
- The CFDE Coordinating Center had "Deep Dive" in-person visits to 8 Common Fund DCCs.
  - **Kids First**
  - **GTE<sub>x</sub>**
  - **HMP**
  - **LINCS**
  - **SPARC**
  - **MoTrPAC**
  - **HuBMAP**
  - **4D Nucleome**
  - **Metabolomics (Nov)**
- Plan to evaluate "success" of the CFDE both impact on CF Programs and project process.
- Fund the next set of CFDE activities, such as addressing data storage and single-sign on.

# How is the Common Fund Addressing Challenges?



The Common Fund

## Common Fund Support



# The Science and Technology Research Infrastructure for Discovery Experimentation, and Sustainability (STRIDES) Initiative



National Institutes of Health  
*Office of Strategic Coordination - The Common Fund*



# What is the STRIDES Initiative?



*The Common Fund*

Through partnerships with commercial cloud providers, the STRIDES Initiative provides a cost-effective way for biomedical researchers at the NIH and at NIH-funded institutions to access rich datasets and the most advanced computational infrastructure, tools, and services provided by commercial cloud providers.



# What does STRIDES provide?

- ✓ Discounts on commercial cloud services (currently Google and Amazon).
  - X It won't create a platform to host or use the data.
- ✓ A mechanism for ICOs to pay for all cloud services centrally, thereby providing a step towards sustainability of data sets after grants end.
- ✓ Access to cloud training for researchers & technical staff.
  - ✓ Training will be offered on-site, at conference/workshop, and online.
- ✓ Access to enterprise-level customer support and help for Programs to learn how to best architect and use each cloud platform.
  - X Programs are responsible for data and account management.
- ✓ Potential collaborative engagements through a community of practice.



Common Fund invested \$16M with the STRIDES Other Transactions Awards divided across two cloud platforms.



Google Cloud



Cloud services will be provided as in-kind services rather than funds to each individual grantee.

# Steps for CF Programs who want to use STRIDES



The Common Fund

Initial STRIDES  
consultation with  
CF Liaison

Develop and submit  
STRIDES  
Recommendation

STRIDES  
Recommendation  
Concurrence

Complete STRIDES  
questionnaire on  
CSP and account(s)

Account(s)  
provisioned

Operational

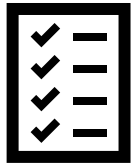
To discuss what the program might need, what STRIDES offers, plan for moving forward

Program Officials and awardees define the goals of the program's STRIDES project, expected budget, etc. The CF/STRIDES team will assist as needed

# CF/STRIDES Recommendation details



The Common Fund

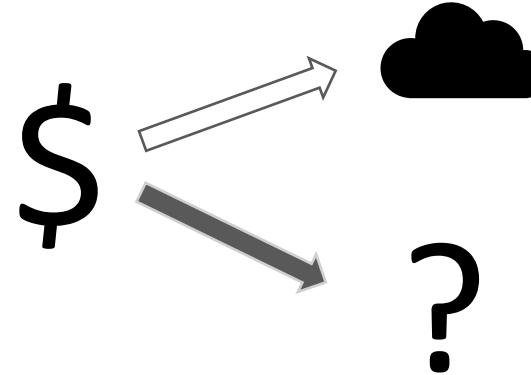


**Program Officials and DCC PIs** will work together to determine costs/needs for cloud services

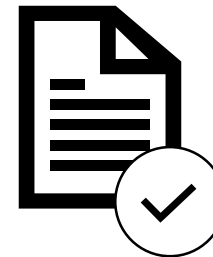


- Data storage
- Compute
- Plans for collaboration with CSP engineers, etc.
- Supporting FAIR

+ assistance from **CF STRIDES Liaisons** as needed



**DCC PIs** will have the latitude to re-purpose funds in their budgets that have previously been provided for cloud costs



**Program Officials and Grants Management** work together to modify the template awards Terms & Conditions for STRIDES

# CFDE

Access to cloud services via STRIDES comes with the expectation that the DCC will work with the CFDE

# STRIDES engagement process for CF Programs



The Common Fund

Initial STRIDES consultation with CF Liaison

To discuss what the program might need, what STRIDES offers, plan for moving forward

Develop and submit STRIDES Recommendation

Program POs and awardees define the goals of the program's STRIDES project, expected budget, etc. The CF/STRIDES team will assist as needed

STRIDES Recommendation Concurrence

CF and the Program come to agreement on the final plan, budget, terms and conditions, etc.

Complete STRIDES questionnaire on CSP and account(s)

Program awardees provide specific information about their needs, the CSPs they wish to use, configuration requirements, etc.

Account(s) provisioned

STRIDES works with the Program and awardees to finalize account details and configure the required account(s) for the Program

Operational

STRIDES provides regular reports on usage and running costs, specific alerts are in place to notify awardees, the program and others at specific spending thresholds

# Which CF Programs are working with STRIDES?



The Common Fund

Program	Had Initial Consultation	STRIDES Recommendation Submitted	STRIDES Recommendation Concurrence	CSP & Account Questionnaire Completed	Account(s) Provisioned	Operational
GTEEx	✓	✓				
HuBMAP	✓					
Kids First	✓	✓	✓	✓		
LINCS	✓					
Metabolomics	✓			✓	✓	
SPARC	✓					

# What do these projects mean for you / your program?



**National Institutes of Health**  
*Office of Strategic Coordination - The Common Fund*



# What is CF asking "my" Program to do?



*The Common Fund*

CFDE Deep Dive  
Site Visit to  
DCCs

Consider  
STRIDES

DCC applies to  
be part of CFDE

Collaboration  
and  
Harmonization

Implementation

More details on these activities and actions for Program Officials and Grants Management are included on supplemental slides 30-35

## For Program Officials

- Look out for FOA
  - Questions to Lora Kutkat
- Talk to PIs about CFDE and STRIDES
- If contacted, work to schedule Deep Dive visit
  - Owen White, Titus Brown, or Amanda Charbonneau
- If interested in STRIDES, reach out to CF-STRIDES Liaison (Simon) to begin conversation
  - Review Common Fund STRIDES Decision Support Tool and Recommendation Document.
  - Work with GM to modify Terms & Conditions

## For Grants Management

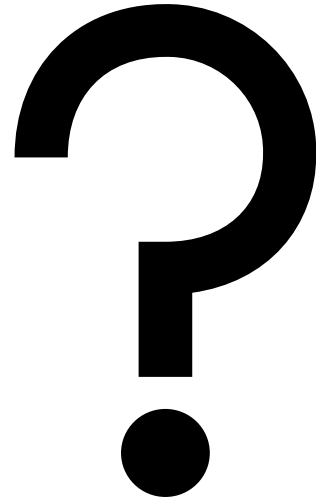
- Talk to Program Officials when FOA released
  - Questions to Lora Kutkat
- If the CF Program is interested in STRIDES:
  - Work with PO to customize the template (OPERA-approved) Terms & Conditions to suit Program.
  - Include the customized Terms & Conditions in the award documents.



# Questions and Discussion



*The Common Fund*



# Supplemental Slides for reference



National Institutes of Health  
*Office of Strategic Coordination - The Common Fund*

# Who to contact



The Common Fund

What	Who	Contact	Role/Contact For
STRIDES engagement	Simon Twigger Todd Reilly Nick Weber	<a href="mailto:simon@bioteam.net">simon@bioteam.net</a> <a href="mailto:todd.reilly@nih.gov">todd.reilly@nih.gov</a> <a href="mailto:nick.weber@nih.gov">nick.weber@nih.gov</a>	Working with CF Programs to leverage STRIDES OTAs, choose a cloud service provider, and establish accounts. Contact to get started on the CF-STRIDES Recommendation Template.
CFDE Coordination Center & DCC Engagement Team	Owen White Titus Brown Amanda Charboneau	<a href="mailto:owhite@som.umaryland.edu">owhite@som.umaryland.edu</a> <a href="mailto:ctbrown@UCDAVIS.EDU">ctbrown@UCDAVIS.EDU</a> <a href="mailto:achar@ucdavis.edu">achar@ucdavis.edu</a>	Building partnerships with DCCs through ongoing engagement with key DCC staff in gathering insights into program goals, issues, and needs. Testing technical standards. Will reach out to Program staff to schedule site visits. Contact with questions regarding standards (e.g. FAIR).
CFDE Project Management	Lora Kutkat Jen Yttri	<a href="mailto:Lora.Kutkat@nih.gov">Lora.Kutkat@nih.gov</a> <a href="mailto:jennifer.yttri@nih.gov">jennifer.yttri@nih.gov</a> <a href="mailto:CFDE@od.nih.gov">CFDE@od.nih.gov</a>	Managing CFDE on behalf of Common Fund. Contact with questions regarding CFDE participation and FOAs. <b>Copy on communications with CFDE Coordination Center.</b>

# Details on: Deep Dive Site Visits to the DCCs



The Common Fund

*In-person visit by the CFDE Coordinating Center to the DCC to understand Program's issues and needs. CFDE wants to understand how researchers use the data and scientific questions that require access to >1 dataset. First step towards establishing a partnership between the CFDE and DCCs.*

## **For Program Officials**

- The CFDE CC invites DCCs to participate. These are in-person with some dial-in by phone.
- DCC PIs will likely ask what this effort is for. Keep DCCs PIs informed about the CFDE Project goals.
- Programs that have not yet been engaged: KOMP2, CryoEM, ExRNA, A2CPS, Undiagnosed Disease Network, Illuminating the Druggable Genome.

## **For Grants Management**

- No Action Needed

# Details on: STRIDES



The Common Fund

*Through STRIDES, CF Programs will be able to leverage discounts for cloud storage and compute on either Google or Amazon and access to training and technical support through in-kind support from Common Fund.*

## For Program Officials

- Review Common Fund STRIDES Decision Support Tool and Recommendation Document.
- Schedule an initial consultation with the CF STRIDES liaison.
- Work with GM to customize template Terms & Conditions to suit your Program.
- Complete the Recommendation Document and submit to OSC.
- Upon approval, work with the STRIDES team to initiate STRIDES billing account creation.
  - Once STRIDES billing account is provisioned, your designated administrator will begin set-up of cloud environment
- Work with the DCC PI to reprogram funding originally dedicated to cloud storage and compute

## For Grants Management

- Work with PO to customize the template (OPERA-approved) Terms & Conditions to suit Program.
- Once program is approved to utilize STRIDES, include the customized Terms & Conditions in the award documents.
- As appropriate, allow for the reprogramming of funds originally dedicated to cloud storage and compute.

# Details on: DCCs Applying for CFDE Funding



The Common Fund

*The Common Fund appreciates the level of effort needed to be part of the CFDE. During FY2020, CF plans to support DCCs as active partners in the establishment of the CFDE. Awards are estimated at \$250K/yr/award, with up to 10 awards in FY2020.*

## For Program Officials

- May be asked to review the Funding/Research Opportunity Announcement.
- May also be asked to review applications.
- Direct any questions about the FOA to Lora Kutkat.

## For Grants Management

- Action needed would depend on whether the award mechanism supplements the current grant award.
- If an Other Transactions award, no action is anticipated from GM.



# Details on: DCC Baseline Survey



The Common Fund

*The Common Fund will ask DCCs that participate in the CFDE to complete a survey in an effort to capture baseline data. A future impact survey will follow in ~2-3 years.*

## For Program Officials

May be asked to:

- Review the survey (some have already provided feedback).
- Confirm survey data that is pre-populated based on the "deep dive."
- Recommend the best contacts to respond to the survey.
- Contribute to CFDE "success" metrics for the impact evaluation.

## For Grants Management

- No action is anticipated.

# Details on: DCC-CFDE Collaboration and Harmonization



The Common Fund

*Once funded, the DCCs will collaborate with the CFDE to make their data sets more useful and usable alone and in combination with other Common Fund datasets. DCCs will be asked to participate in activities such as creating standards for data management, testing single-sign-on, and sharing data.*

## **For Program Officials**

Will be invited to:

- Attend DCC-to-DCC workshops.
- Provide support to DCCs as they participate in joint exercises and other activities.
- Review the success of joint exercises.
- Contribute to use cases and end-user training.
- Offer technical insights.
- Offer feedback on the CFDE process.

## **For Grants Management**

- No action is anticipated.

# Details on: Implementation



The Common Fund

*After initial collaboration and harmonization efforts, DCCs engage in longer term efforts to address long term data sustainability. This will include establishing community determined standards, demonstrating interoperability, and engaging users to perform new scientific discoveries.*

## For Program Officials

May be invited to:

- Test their own use cases for data interoperability.
- Discuss options for data sustainability after a Common Fund Program ends.
- Participate in a process evaluation.
- Develop follow-on Funding Opportunity Announcements that utilize multiple Common Fund datasets.

## For Grants Management

- No action is anticipated