CoC Working Group on the Common Fund Data Ecosystem (CFDE)

Final Report Presentation

Betsy Wilder and Rick Horwitz
May 20, 2022

NIH National Institutes of Health
Office of Strategic Coordination–The Common Fund

Common Fund Programs Catalyze Biomedical Discovery

Common Fund makes substantial investments in time-limited, goal-driven programs that significantly change the trajectory of biomedical research.

Accelerate emerging science

Remove research roadblocks

Enhance the research workforce
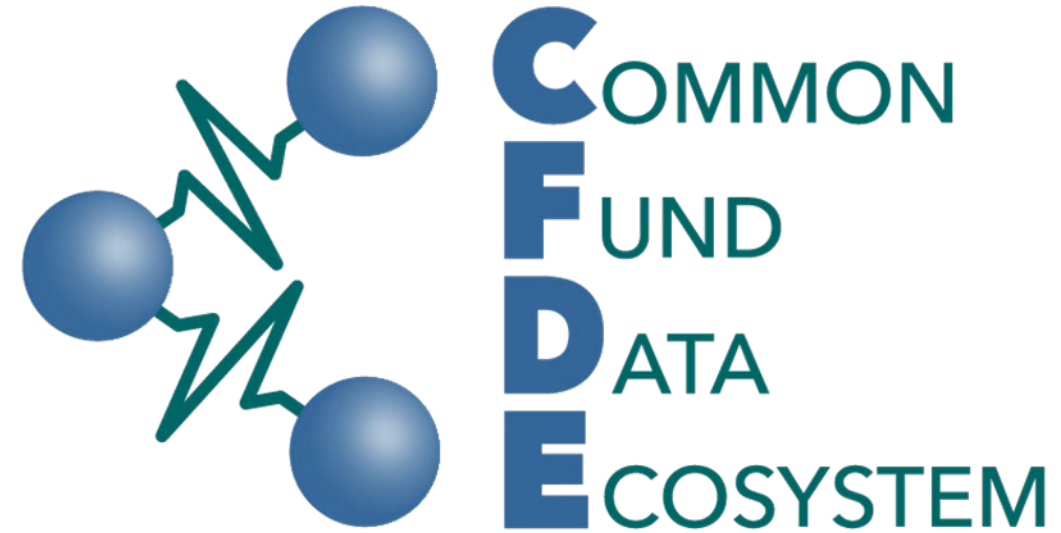
Support high-risk, high-reward science

CF programs often develop broadly useful resources, including large data sets and computational tools

The Common Fund Data Ecosystem (CFDE)

3 Motivating Questions:

- How would a researcher **combine data sets** from multiple Common Fund programs?

- What happens to Common Fund data **after** the program's **funding** ends?

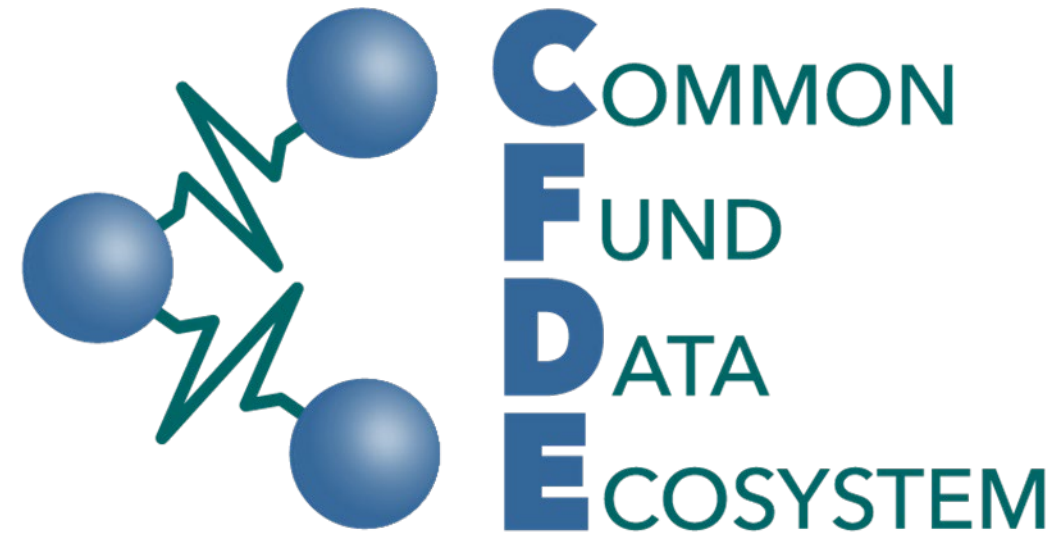- With data moving to the cloud, how can we **train** users to work with the data there?

COMMON
FUND
DATA
ECOSYSTEM

CFDE Vision and Goals

Goal: create a functioning ecosystem that makes Common Fund (CF) data sets more impactful by enabling easier data access and use

CFDE aims to:

1. Enable users to **query across & use** multiple CF data sets

2. **Sustain** CF data and tools

3. **Train** users to work with CF data

COMMON
FUND
DATA
ECOSYSTEM

CFDE is intended to support and enhance the value of many Common Fund programs
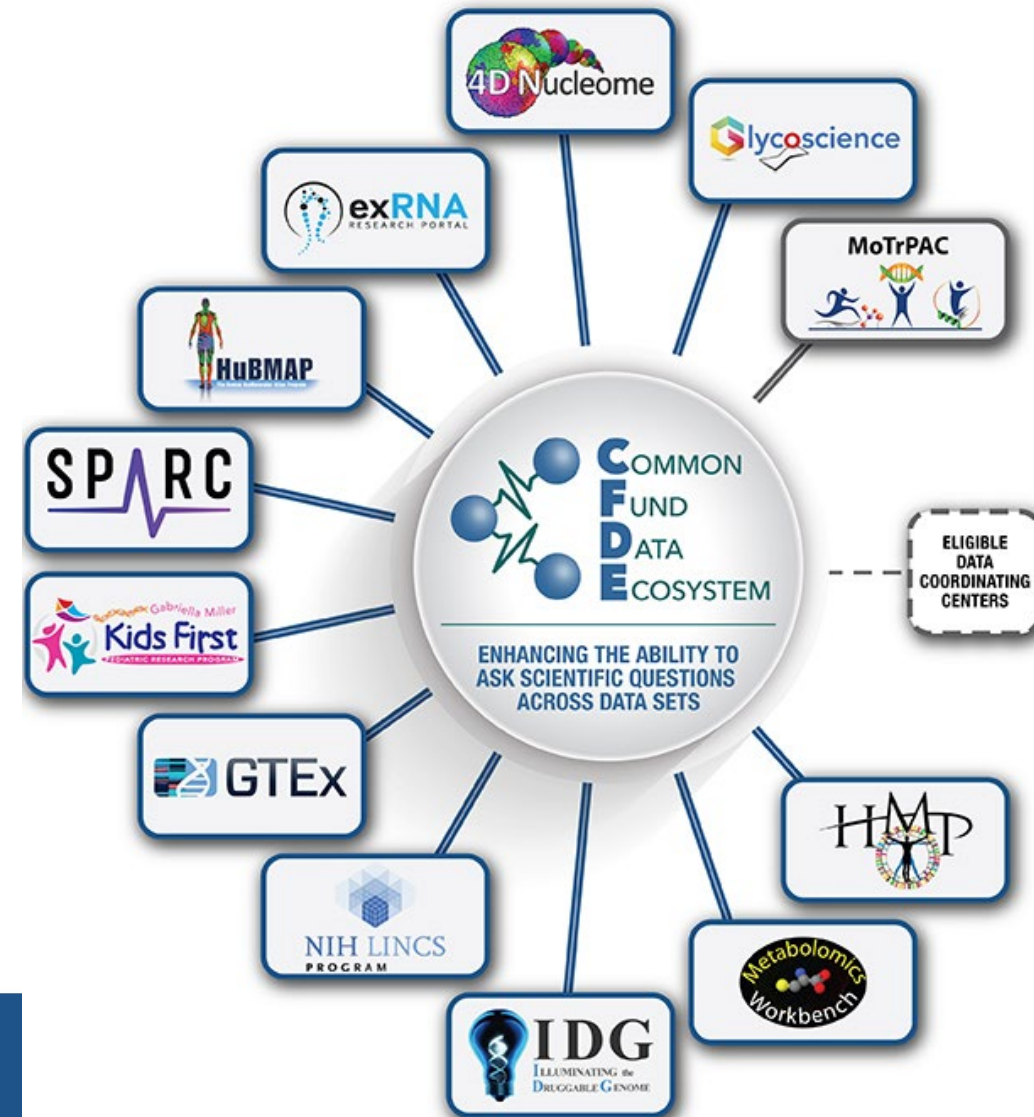
Where is CFDE Today?

CFDE Coordinating Center & 11 CF programs Data Coordinating Centers (DCCs) are **building the data ecosystem**

- CFDE Portal to find relevant data sets by searching across metadata
- DCC joint projects exploring scientific questions addressed through interrogation of multiple data sets

22 research awards (R03s) to **enhance utility and use of CF data sets**

Developing an **assessment** plan to measure progress & impact of the program

CFDE CoC Working Group

Working Group Charge

Review the current scope and goals of the CFDE as well as progress to date

Make recommendations for future scope and goals in the following areas:

- Findability and accessibility of data
- Data harmonization and interoperability
- Cloud workspaces
- CFDE scope and strategy in the context of related NIH activities

Interoperability and Reuse of Data

- Sustaining access to data and tools after CF programs end

Sustainability and Accessibility

- Training and outreach to enhance access to, and use of, the data

Training

CFDE CoC Working Group Members and Activities

Working Group Members

Rick Horwitz, Ph.D. (Chair)
Allen Institute for Cell Science

Sergio Baranzini, Ph.D.
University of California, San Francisco

Emiley Eloe-Fadrosh, Ph.D.
Lawrence Berkeley National Laboratory

Warren Kibbe, Ph.D.
Duke University School of Medicine

Elizabeth Wilder, Ph.D. (Chair)
OSC, DPCPSI, NIH

Devin Schweppe, Ph.D.
University of Washington

Bruce Weir, Ph.D.
University of Washington

Cathy Wu, Ph.D.
University of Delaware

Working Group Activities

- Monthly meetings, July 2021 – April 2022 (all virtual)
- Presentations from NIH data science experts and CFDE awardees
- Discussions with presenters, closed discussions with Working Group
- Meeting topics included: status of CFDE, harmonization/interoperability, cloud workspaces, training, sustainability, NIH data science activities

CFDE CoC Working Group – Key Observations

The challenges that CFDE faces with respect to managing data resources are shared by data managers throughout the biomedical research enterprise. CFDE will not solve these challenges alone, and some challenges are beyond the scope of CFDE.

The pilot phase of the CFDE has been impressive and accomplished much in three years, including development of a discovery portal, establishment of a metadata model, piloting a cloud-based workspace, training for users, support for exploration across data sets, and establishment of a community of investigators who are collaborating to build the ecosystem.

The chief metric of success for CFDE is discovery: if the CFDE is successful, in five years many investigators will be using Common Fund data for new discoveries and new purposes.

CFDE CoC Working Group – Synthesis of Recommendations

Interoperability and Reuse of Data

- Support queries across data sets through:
  - **Use of community-based standards**
  - **Development of robust use cases and workflows for diverse data users**
  - Development of innovative methods for data search
  - Expanded use of knowledge graphs that provide relationships across data sets
  - Citation of data/tool contributors by CFDE users

- User experiences should shape further development of CFDE

- CFDE should actively evolve through development of tools for new uses/users and making data sets available in the cloud

- Continued attention to the NIH-wide and other data communities is essential

National Institutes of Health
Office of Strategic Coordination–The Common Fund

CFDE CoC Working Group – Synthesis of Recommendations

Sustainability and Accessibility

- Public repositories offer an increasingly attractive option for sustaining data beyond Common Fund support

- Repositories developed by Common Fund program should continue to transition to support from NIH Institutes and Centers or other entities, if they prove useful to the community

- CFDE should not be a long-term repository for Common Fund data sets

- Repository and data management are NIH-wide concerns beyond the scope of CFDE; CFDE should participate in these efforts as an important client of these repositories

- Cloud storage may be an option, but barriers to democratization need to be considered

CFDE CoC Working Group – Synthesis of Recommendations

Training

- Training users needs to be an enhanced priority for CFDE

- Focus of training should be on enabling use of Common Fund data sets

- CFDE should consider diverse training approaches, including an emphasis on scalable training (e.g., videos, MOOCs, FAQs, help desks)

- Individualized training efforts (fellowships, courses, small grants, etc.) are also valuable and should emphasize early-stage investigators and investigators from lower-resourced institutions

Council Action:
Vote to accept the Working Group Report on CFDE

🌐 commonfund.nih.gov

f @NIHCommonFund

🐦 @NIH_CommonFund

NIH〉 National Institutes of Health
Office of Strategic Coordination—The Common Fund