

Proposal for Phase 2 of the Common Fund Data Ecosystem (CFDE)

CFDE's driving focus is to foster scientific discovery through (re)use of data generated by the Common Fund (CF) programs. This will be achieved through three aims: 1) Enable users to query across & use multiple CF data sets, 2) Provide training and outreach to bring people to CF data and train them to work in the cloud, and 3) Coordinate and integrate infrastructure and activities into a cohesive ecosystem.

CFDE Phase 1 Progress: The major accomplishments of CFDE's first phase (2019 – 2022) include establishing an initial [discovery portal](#), where users can search for datasets and resources across CF programs; prototyping knowledge graphs, e.g., [ReproTox Knowledge Graph](#); piloting of a CFDE focused [cloud workspace](#) capable of receiving data from the CFDE portal for analysis with a user's data; creating [training materials](#) to help promote the uptake of cloud resources; and establishing a community of investigators who are collaborating to build the ecosystem. Additional video-based highlights are available [here](#).

CFDE Phase 2 Program Goal: CFDE's Phase 2 will continue to build on the Phase 1 success by pursuing its three aims through increased pacing and expansion. CFDE will be considered a success if many investigators use CF data for new discoveries and purposes.

Initiatives 1 and 2. Data and Knowledge Portals: The Data Resource Portal will catalog CF data sets and CFDE resources in a user-friendly portal, incorporating state-of-the-art search capability, enabling users to find CF data sets and resources and then import those datasets into a cloud workspace(s). Similarly, the Knowledge Portal will build the infrastructure needed to integrate knowledge generated by the CF programs, connect it to the complementary external knowledge resources, and develop a mechanism to ensure it can be used in cloud workspace(s).

Initiative 3. Cloud Workspace: The Cloud Workspace initiative will create the resources needed for both novice and expert users to utilize CF data and resources in the cloud. For novice users, it will quickly form a curated multi-omics workspace, providing a helpdesk and well-documented, interface-driven tools to analyze CFDE data. For experienced users, it will create the "CFDE Workbench" to integrate CF and cloud-based datasets with community and CFDE-developed "dockerized" tools.

Initiative 4. Training and Outreach to be achieved via 5 sub-initiatives: The Center for Training will create a syllabus for scalable online training that will initially focus on cloud computing and transition to bioinformatics/multi-omics over time as cloud skills become integrated. Sub-initiatives to support diversity-focused Summer Courses, two-year pre- and post-doctoral diversity fellowships, and diversity supplements will aim to expand the Common Fund data user community and provide opportunities for early-stage investigators to build research programs using data-intensive approaches. A fifth sub-initiative continues an existing small grant program to draw attention to Common Fund data sets and encourage any investigator to develop new lines of research using these data.

Initiative 5. Integration and Coordination Center: The Integration and Coordination Center (ICC) will build cohesion within CFDE by integrating relevant CF programs within CFDE, coordinating across the centers, helping CF programs develop data sustainability plans, and ensuring continuous improvement through an annual evaluation process.

Initiative 6. Data Coordination Center Engagement: Data Coordinating Centers (DCCs) of the engaged CF programs participate in infrastructure and tool development for integrative data analysis and knowledge generation. CFDE will continue to support these activities to foster harmonization, interoperability and increased reuse of the CF-generated data and resources to maximize the impact of the CF programs.

Proposed Budget (in millions)

Initiative	Title	FY23	FY24	FY25	FY26	FY27	Total
1	Data Portal	\$1.75	\$1.75	\$1.75	\$1.5	\$1.25	\$8
2	Knowledge Portal	\$1.25	\$1.5	\$1.75	\$2	\$2.25	\$8.75
3	Workspace	\$0.75	\$1.5	\$1.5	\$1.5	\$1.5	\$6.75
4	Training and Outreach	\$7	\$7.9	\$7.7	\$7.7	\$7.1	\$37.4
5	Integration and Coordination	\$1.5	\$2.25	\$2	\$2	\$2	\$9.75
6	DCC Engagement	\$8	\$8	\$8	\$8	\$8	\$40
N/A	RMS	\$0.2	\$0.2	\$0.2	\$0.2	\$0.2	\$1
Total	N/A	\$20.45	\$23.1	\$22.9	\$22.9	\$22.3	\$111.65