Concept Clearance for the Common Fund Data Ecosystem (CFDE)
Funding Opportunity: Targeted Needs at CF Data Coordinating Centers to Establish the CFDE

As described at the May, 2019 Council of Councils meeting, the overarching goal of the CFDE is to extend from and leverage deliverables from the Data Commons Pilot Phase Consortium to enhance utility of Common Fund Data Sets. This will involve: 1) making CF data sets more useful/usable within a program and between programs by improving Findability, Accessibility, Interoperability, and Reusability (FAIRness), 2) capturing and developing best practices for new programs to leverage, 3) enhancing the ability to ask scientific questions across data sets, 4) increasing reuse of data and tools after a program ends, and 5) providing a pathway to sustainability after CF programs end. These goals support the goals of the Trans-NIH Strategic Plan for Data Science and are being addressed in coordination with the NIH Office of Data Science Strategy and NIH ICs.

The CFDE is thus intended to provide ongoing data management infrastructure that will support past, existing, and future CF datasets. For new programs, the CFDE will provide guidelines and best practices to foster FAIRness. For existing programs, it will provide a forum for the articulation and adoption of best practices and for sharing information. For programs that are ending or that have ended, it provides a pathway to sustainability for CF data sets by providing basic services that will continue to make the data available. Therefore, the individual Data Coordinating Centers (DCCs) that the CFDE supports will turn over in line with the expected 10 year time frame of Common Fund programs, but the CFDE itself will be sustained through ongoing support of a CFDE Coordinating Center (CFDECC).

The CFDECC, which was established through the Data Commons Pilot Consortium but now has re-focused on CF programs, will support the CF (DCCs) by working with each DCC to understand their specific requirements, by providing a forum for information exchange between DCCs, by developing training modules for user communities, and by developing and implementing sustainability plans when programs end. Long term sustainability of data will depend on user demand, so a key function of the CFDECC will be to assess the utility of each of the data sets within the CFDE.

In addition to the CFDECC, the CFDE will also require active engagement by CF DCCs. To date, the CFDECC has had "deep dive" engagements with four DCCs to outline a set of activities that will be required to establish the CFDE. It has become clear that each DCC will need to dedicate specific effort and resources to participate in CFDE meetings, to establish or adapt and adopt CFDE FAIR guidelines and best practices, to mentor/share information with staff from new DCCs, to establish use cases for cross-data queries and enable those analyses to be conducted, to help users operate in a cloud environment, and thus to help expand the science that can be conducted and speed the use of data in translational and clinical applications.

This concept is therefore a proposal for targeted, limited competition solicitation of applications from CF DCCs to engage with the CFDECC and other DCCs to establish the CFDE. The initial duration of awards through this solicitation will be for 3 years. We expect to re-solicit applications as DCCs for new programs are established. Budgets are expected to vary depending on the specific requirements of each DCC, but are estimated at $250K/year/award.