

## **Title of proposed program:** Coordinating Data Collection

**Submitting Source:** Strategic Planning Meetings

### **What is the major obstacle/challenge/opportunity that the Common Fund should address?**

Large sample sizes increase statistical power and can lead to greater reliability of research results, whereas comparable studies which rely on small samples often yield contradictory findings. Greater coordination and cooperation among investigators in collecting and pooling the data from studies has the potential to significantly increase knowledge in all fields. The development of a culture of collegiality among investigators coupled with technological advances in the storage and access to pooled data could break down barriers between labs and allow investigators greater flexibility to publish more reliable findings. For this model to be successful, however, issues of privacy with regard to data on individual human subjects and the use of these data in research would need to be addressed.

### **What would the goals of the program be?**

The program would have 6 goals:

- Establish procedures to collect standardized data from diverse sets of investigators
- Standardize phenotypes
- Remove barriers to capturing and sharing data as a consortium when groups did not initially develop as consortia
- Expand the data pool to include school districts, state health care systems, and similar entities
- Establish interoperability standards and a data warehouse for data mining and sharing data
- Address the issues and ethics surrounding privacy

### **Why is a trans-NIH strategy needed to achieve these goals?**

The cultural, technological, and privacy issues that form barriers to data coordination transcend the boundaries that separate the missions of the various ICs. These issues are relevant to data collection across disciplines and require the attention all parts of NIH.

### **What initiatives might form the strategic plan for this topic?**

Goals that included a discussion of initiatives included the following:

- Establish procedures to collect standardized data from diverse sets of investigators
  - Develop a list of different issues that require an integrated approach
  - Collect data from each human subject for use in studies across a variety of different domains
- Remove barriers to capturing and sharing data as a consortium when groups did not initially develop as consortia
  - Incentivize participation by investigators
- Establish interoperability standards and a data warehouse for data mining and sharing data
  - Build an NIH-housed data repository, analogous to PubMed, which is accessible to investigators both to input and access data
- Address the issues and ethics surrounding privacy
  - Create a stipulation in consent forms about the sharing of data in the data repository

### **If a Common Fund program on this topic achieved its objectives, what would be the impact?**

A successful program to coordinate data collection could lead to more robust findings based on larger sample sizes, an outcome which will contribute to more rapid advancement across the fields of health. It could potentially standardize both the definitions and measures of variables of interest and create an infrastructure to capture findings from multiple studies in a manner that makes data accessible while it incentivizes collegiality among investigators and fosters collaboration between disciplines.