

## **Concept Clearance - Reissue of the Predoctoral Training in Advanced Data Analytics for Behavioral and Social Sciences Research (BSSR) - Institutional Research Training Program [T32]**

### **Description:**

In 2020, OBSSR noted that although the behavioral and social sciences had seen an explosion of data sources from smartphones, sensor technologies, electronic health records, social media, and administrative data repositories, graduate programs continue to train our doctoral students in early methods of data science. Thus, in partnership with the National Heart, Lung, and Blood Institute, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute on Minority Health and Health Disparities, OBSSR launched a 5-year training program to incorporate computational and data science analytic approaches directly into behavioral and social sciences predoctoral degree programs. The goal was to support the development of a cohort of specialized scholars pursuing careers in health-related research with competencies in data science analytics.

Current training program titles include:

- Data Integration for Causal Inference in Behavioral Health;
- Data Science Training in Demography and Population Health; and
- Training in Advanced Data Analytics to End Drug-Related Harms.

This concept clearance proposes to extend this successful program for an additional five-year competitive cycle. In summary, predoctoral students are expected to acquire core knowledge in three overarching relevant areas in:

1. Computer science/informatics [data mining, pattern recognition, machine learning computational modeling, natural language processing];
2. Statistics/mathematics; and
3. Behavioral or social sciences research in a chosen health domain relevant to the NIH Institutes and Centers (ICs) participating in this program announcement.

### **Phase 1 Accomplishments:**

The current cohort of this 5-year training program includes 8 OBSSR-funded awards made in FY 2020 that are administered by 5 ICs. A quantitative and qualitative evaluation is currently underway of this program and results will be forthcoming in early FY 2024. To date, this program has supported 64 trainees (54 current and 10 alumni) across diverse disciplines including sociology, epidemiology, public health, health economics, health systems and services, mental health, and public policy. Trainees who have completed their graduate programs have moved into related academic fields of study (in positions such as post-doctoral fellow and research associate). Trainees have authored or co-authored 85 publications to date in peer-reviewed scientific journals. Trainees and investigators have also secured additional funding in the form of diversity supplements and a fellowship.

### **Phase 2 Concept:**

To continue to build a cadre of predoctoral BSSR candidates, OBSSR requests approval to reissue this Notice of Funding Opportunity. Specific areas of interest include:

- Intensive or voluminous longitudinal data;
- Electronic data (electronic health records (EHR), social media, commercial, and administrative records data; and
- High-density, large sample or population-level agency databases.

### **Funding Information and Timeline:**

The total estimated cost for this initiative is \$12M over 5 years. The current competitive cycle makes its final awards in FY 2024. This next cycle will cover FY 2025-29. We anticipate another round of eight awards of similar size to the current cohort. The competition will be open to both existing and potential new awardees.