**Cybertraining: Pilot: SBE: Training Modules for Hypothesis-Free Inquiry on Supercomputing Platforms**

This project will develop an integrated suite of training modules to educate public health researchers and ancillary social scientists in the methods and practice of hypothesis-free discovery on supercomputing platforms. The project outcomes will enhance the technical capabilities of interdisciplinary researchers conducting environment-wide association studies (EWAS). Typically, investigators engaged in these kinds of studies are seeking to better understand the economic, social, environmental, and biological factors that contribute to health and disease in various populations.

In an age of machine learning, big data, and supercomputing, a new scientific method is emerging. Instead of formulating a hypothesis, executing an experiment, and then drawing a conclusion, a hypothesis-free approach to scientific inquiry is now possible. In this new world - made possible by advanced cyberinfrastructure - a scientist finds and then links multiple datasets, followed by the work of an algorithm (usually a neural network) which "discovers" possible correlations between variables. Afterwards, the scientist reviews the "discoveries" made by the software and validates their reasonableness, discarding irrelevant relationships.

1. Data discovery: This is a highly creative step in which the researcher finds datasets to place in juxtaposition to each other.
2. Data cleaning : This step prepares the data for linking.
3. Data linking : Disparate datasets are linked, allowing significant relationships to be examined by a neural network.
4. Machine learn : A neural network is trained to discover correlations/relationships in the data.
5. Verification : Potential relationships discovered by the neural network are validated and/or discarded.

SBE welcomes proposals that address research training challenges, such as educating SBE scientists for understanding and promoting economic opportunity, security, civic and political engagement, health, and well-being in different regions and populations, many of which require interdisciplinary sociotechnical collaborations and team science.

**SBE** supports rigorous methods to discover fundamental principles of human behavior at levels ranging from cells to society, from neurons to neighborhoods, and across space and time. The SBE directorate supports research that advances computational social science and analytic methods using social network, sensor, text, video, administrative, and other big data. SBE seeks proposals that will advance training for research communities of CI Professionals and CI Users. SBE welcomes proposals that address research training challenges, such as educating SBE scientists for understanding and promoting economic opportunity, security, civic and political engagement, health, and well-being in different regions and populations, many of which require interdisciplinary sociotechnical collaborations and team science.

**Cover Sheet:**The projects will have a short informative title that begins with “CyberTraining: Pilot:”, “CyberTraining: Implementation: Small:”, “CyberTraining: Implementation: Medium:”, or “CyberTraining: Conceptualization:” based on the project class chosen.

**Project Summary (1-page limit):**The Project Summary consists of an overview, a statement on the intellectual merit of the proposed activity, and a statement on the broader impacts of the proposed activity. The overview includes a summary description of the project, including the need for the activity and its short- and long-term goals for training, education, and research workforce development; broadening CI access and adoption goals, or curriculum/instructional material goals; innovative aspects; size and nature of target CI communities; and estimated number of students. The overview also includes **6-7 keywords** that specify the CI community (i.e., *Professionals, Contributors, Users*), level of students, disciplines, topics, and themes targeted. The Project Summary should be written in a manner that will be informative to STEM professionals working in the same or related fields, and understandable to a scientifically-literate lay reader.

**Project Description (15-page limit):**The project description should explicitly address a subset of the following additional items, according to the project class, with emphasis suitable to the proposed work and chosen goal(s) of the solicitation (note that this information will also be employed as additional solicitation-specific review criteria; see Section VI.A. for details):

1. Challenges for Research Workforce Development;
2. Solicitation Goal(s) Targeted (at least one goal for*Pilot* and *Implementation* proposals, and both goals for *Large-scale Conceptualization* Project proposals):
   1. Broadening Adoption of Advanced CI; or
   2. Integration of CI Skills into Curriculum/Instructional Material Fabric;
3. Scalability and Sustainability;
4. Recruitment and Evaluation;
5. “Collective Impact” Strategy: Coordination network and Backbone organization (or an alternative strategy);
6. Fostering Community;
7. Information Hub and Repository Infrastructure; and
8. Support and Guide for other CyberTraining and relevant projects, and the community.

*Pilot* project must address items 1 and 2. *Small Implementation* projects must address items 1-5, and *Medium* *Implementation* projects must address items 1-6. *Large-scale Project Conceptualization*projects must address all 8 items, including both solicitation goals.

Please note that, per guidance in the PAPPG, the Project Description must contain separate sections labeled "Intellectual Merit" and "Broader Impacts."