Concept Clearance

Branch: GEHB, ERTB, HSRB

Council Period: 202001

Concept Title: NIEHS Program to Support Training & Workforce Development in Environmental Health Data Science

Introduction

With rapidly developing technology and more efficient data collection procedures, biomedical scientists are now collecting vast amounts of data, and modern biomedical research is becoming increasingly quantitative. The National Institutes of Health (NIH) considers it essential to equip the next generation of researchers with the skills needed to take advantage of the growing promise of data science for advancing human health. Accordingly, an overarching goal of the NIH Strategic Plan for Data Science is to enhance workforce development for biomedical data science.

Within the biomedical realm, Environmental Health Sciences (EHS) data are among the most diverse, complex, and challenging. EHS research encompasses the study of all levels of biological organization – molecular, biochemical pathway, cellular, tissue, organ, system, model organism, individual, and population – at all stages across the lifespan, from preconception through old age. EHS uses a rich, diverse, and constantly evolving set of observational, experimental, computational, and clinical approaches to explore the impacts of varying levels of exposure and resilience/susceptibility to such exposure on human health. EHS data types include high-dimensional data (e.g., genomic, metabolomic), demographic, questionnaire, qualitative, electronic health record, imaging, geospatial, and toxicological data. The National Institute of Environmental Health Sciences (NIEHS) recognizes that appropriate tools, access to quality datasets, and training are needed to harvest the wealth of information contained in these data to advance our understanding of the role of the environment in human health and disease. Correspondingly, a focus of the NIEHS 2018-2023 Strategic Plan is to build a workforce that is both highly qualified in the use of cutting-edge knowledge, technologies, and approaches and dedicated to applying them to solve environmental health problems.

NIH made early efforts to address the gap between the needed and existing biomedical data science skills through investments in training and education as part of the Big Data to Knowledge (BD2K) Initiative. These NIH-wide training efforts, led in part by NIEHS, were not domain-specific and were intended to develop resources which could benefit all NIH institutes. Additionally, a number of other collaborative research and training efforts, including existing quantitative and EHS training programs, have started bridging research and educational efforts in biomedical science and data science. However, significant challenges remain in building an EHS workforce that is fully poised to capitalize on advances in data science and information technology.

In August 2018, the NIEHS convened an interdisciplinary workshop to explore strategies to develop a data science competent EHS workforce. The workshop brought together experts from relevant research disciplines to examine existing data science and EHS resources (including trainee pipelines, mentors, and research) and identify how these resources can address EHS-specific training goals in data science. Additionally, NIEHS collected feedback from EHS trainees on their experiences with big data and data science training, including their most valuable quantitative skillsets and prior training experiences, largest hurdles to gaining new skillsets, and specific training or career development needs to enhance data science skillsets. NIEHS also gathered input from NIEHS Training Program Directors on current approaches for incorporating data science into EHS training programs, as well as the challenges or barriers to meeting the growing need to provide trainees with opportunities to enhance their data science skillsets.

The broad input collected from the EHS and data science communities drew attention to current gaps and barriers to advancing data science research and training. Paralleling the diversity of science in the EHS field, data science training needs for the EHS workforce broadly span scientific expertise areas as well as the full spectrum of career stages. As training needs vary greatly based on an individual's prior background and intended use for data science knowledge and skills, a variety of training opportunities are necessary. Furthermore, there is a growing need to foster partnerships between health researchers and quantitative scientists, improve access to EHS data to drive innovation and training opportunities, and support creation or adaptation of open educational resources and short courses at the intersection of EHS and data science. NIEHS can adapt and extend existing models and strategies developed through BD2K and other quantitative programs. However, the field of data science is fast-moving, and NIEHS recognizes the need for training goals and approaches to evolve in conjunction with advancing technologies.

Research Goals and Scope

NIEHS is proposing a broad, multi-phased concept to advance workforce development for environmental health data science. The proposed program has two main goals:
1. To improve the data science skills of all environmental health scientists.
2. To increase the number of environmental health data scientists, including (a) quantitative-trained researchers to develop tools and methods for environmental health and (b) researchers employing advanced applications of data science for environmental health.

Mechanism and Justification

To achieve the goals of the program, NIEHS is proposing to employ a phased, multi-component approach. Based on input from the EHS and data science communities, guiding principles for this program include leveraging existing resources and adapting existing training models, embracing diverse training approaches to support multiple pathways, recognizing and fostering the interconnection of training and the scientific research base, facilitating cross-domain communication, collaboration, and team science, and maintaining agility in a fast-moving field. Additionally, in order to have maximal impact, it is crucial that training and educational resources developed through this program be Findable, Accessible, Interoperable, and Reusable (FAIR).

Tier 1 of the program consists of two funding initiatives:

1. <u>Administrative Supplements to NIEHS Research Grants to Support Training of Data Scientists in the Environmental Health Sciences Domain</u>. This proposed supplement program provides funds to support research and training experiences of exceptional data scientists. Supplements are intended to support the training of a quantitative scientist who can develop novel tools, methods, or technology to apply to the data and biomedical question being explored in the parent NIEHS research grant. The program is modeled after the existing Research Supplements to Promote Diversity in Health-Related Research Program and focuses on the training and education of the candidate quantitative scientist, who should use the supplement to expand their expertise in the biomedical area of the parent grant. It is anticipated that administrative supplements may be awarded as early as FY21.
2. <u>Open Educational Resources for Environmental Health Data Science</u>. This proposed educational resource program is intended to support innovative data science-focused research educational activities that complement and/or enhance training of the broad EHS workforce, engage data scientists in EHS research, and/or foster formulation of transdisciplinary teams to solve data-intensive environmental health research questions. Educational resources supported by this program may include support for in-person or electronic short courses designed to develop technical, operational, and/or professional skills necessary to conduct data science-focused environmental health research. Dissemination of educational materials and outreach activities openly to individuals from a variety of backgrounds is a required component of the program. The program is modeled in part after the NIGMS Innovative Programs to Enhance Research Training (IPERT) Program and utilizes the R25 mechanism. It is anticipated that Open Educational Resource grants may be awarded as early as FY22.

Future Tiers of this program will adapt according to evolving needs of the field and may include support for Administrative Supplements to NIEHS Institutional Training Programs, Career Development or Career Transition Awards in Environmental Health Data Science, Targeted Fellowships in Environmental Health Data Science, and/or support for an Environmental Health Data Science Training Coordinating Center.

When implemented, these funding opportunities will promote skill development across the broad EHS domain and advance the establishment of data science career paths in EHS. Moving forward, individuals and teams with data science expertise are in a unique position to make lasting impacts on the EHS field. Anticipated long-term impacts of this program include the enabling of novel discoveries that capitalize on advances in data science and benefit the overall EHS research portfolio. Short-term metrics of success include increased numbers of data science-focused EHS grant applications and publications, the development and/or application of novel tools and methods, as well as increased numbers of transdisciplinary collaborations.