TRANSDISCIPLINARY RESEARCH IN PRINCIPLES OF DATA SCIENCE (TRIPODS)

WEBINAR

November 15, 2016

https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505347
Welcome from NSF Assistant Directors James Kurose (CISE) & Fleming Crim (MPS)
Division Directors Rao Kosaraju (CCF) & Michael Vogelius (DMS)
Overview of Program: Program Directors Nandini Kannan (DMS) & Tracy Kimbrel (CCF)
Solicitation Specific Requirements
TRIPODS Review Criteria
Q&A: +Jack Snoeyink (CCF), Chris Stark (DMS), Chaitan Baru (CISE)
OVERVIEW

- Collaboration between the Division of Computing and Communication Foundations (CCF) in the Directorate for Computer & Information Science & Engineering (CISE) and the Division of Mathematical Sciences (DMS) in the Directorate for Mathematical and Physical Sciences (MPS)
- Addresses one of NSF’s 10 Big Ideas “Harnessing Data for 21st Century Science and Engineering”
- Focuses on the theoretical foundations of data science, core algorithmic, mathematical, and statistical principles.
NEED FOR TRIPODS

- NSF funded workshop TFoDS, “Theoretical Foundations of Data Science: Algorithmic, Mathematical, and Statistical,” with researchers representing the three core disciplines of computer science, mathematics, and statistics.

- Key conclusions from the report:
  - “Theoretical foundations are necessary in all aspects of data science, from the generation and collection of data to the analysis and decision making processes.”
  - “Data science is intrinsically interdisciplinary, in the sense that many different scientific domains will need to work together and develop novel theories that transcend disciplinary boundaries.”
Transdisciplinary Research In Principles Of Data Science (TRIPODS) aims to bring together communities from statistics, mathematics, and theoretical computer science to develop the theoretical foundations of data science through institutes for integrated research and training activities.
Supports the development of small collaborative Institutes that will bring together the three disciplines.

Proposals must address fundamental research and training in the theoretical foundations of data science, and describe the significant involvement of all three communities.

Approximately 8-10 awards expected
Teams will develop capacity and demonstrate the ability to scale activities for full Institute operations by operating as smaller Institutes.

- Novel approaches encouraged
- Traditional center-like activities are anticipated
  - workshops
  - training of students & post docs
  - workforce development
  - community building
Encourage PIs to leverage existing NSF investments as appropriate through collaborations or partnerships
- Big Data Regional Innovation Hubs
- Mathematical Sciences Research Institutes
- Software Infrastructure for Sustained Innovation

TRIPODS Phase II: smaller number of larger Institutes, selected from the Phase I Institutes
- via a second competitive proposal process
- to be described in an anticipated future solicitation
- subject to availability of funds
BROAD THEMES OF THE PROGRAM

- Overcoming barriers related to different terminology and formalisms for overlapping concepts and methods developed by different communities
- Relevance to application domains and industry
- Unified curricula for data science: addressing experimental validation, ethical behavior, and interdisciplinary communication skills
- Addressing the entire "data to knowledge to action" pipeline, including dynamic data collection
POSSIBLE RESEARCH FOCI

- Combinatorial inference on complex structures
- Tradeoffs between computational costs and statistical efficiency
- Randomized numerical linear algebra
- Representation theory and non-commutative harmonic analysis
- Topological data analysis (TDA) & homological algebra
- Machine learning including deep learning

Projects must exemplify the roles of the three communities in laying theoretical foundations for data science, and bring about a true synergy of the best capabilities of all three disciplines.
PROPOSAL PREPARATION & REVIEW CRITERIA
Required Elements (in addition to the requirements listed in the NSF PAPPG)

- **Letter of intent required**
  - Submission window: January 4-19, 2017
- **Full proposals**: March 1-15, 2017
- **Collaboration and Evaluation Plan:**
  - Describe expertise in the three disciplines
  - Plans for working together to meet the goals of the program
  - Clear measures of success for the project, including developing capability and capacity for a potential Phase II
  - Plans for evaluating success
Project Description:

- Include an overall **timeline** of activities and **milestones**
- Indicate how the Institute will develop **capacity and capability** for full operations in a potential Phase II

Single Copy Documents: Collaborators & Other Affiliations information specified in the PAPPG should be submitted using the spreadsheet template found at https://www.nsf.gov/cise/collab/
Well-integrated research and training program focused on the theoretical foundations of data science and fostering collaboration and interaction among the three communities of **TRIPODS** – mathematics, statistics, and theoretical computer science.

“**Broad themes of the program**” listed in the Program Description.

Strategies for **workforce development**, including novel educational and training activities.
Transdisciplinarity/Synergy:
- Bringing together theories and approaches from theoretical computer science, mathematics, and statistics
- Synergy between the groups

Vision: Ability to identify and articulate a vision for the foundations of data science

Quality and Value of Collaboration:
- Project expertise is complementary, and well-suited to the research and training programs
- Specific roles of each collaborating investigator are made clear
- Collective team’s expertise represents the three communities

Well-developed plan for communication and interaction with the domain sciences and industry
Clear plan and rationale for an investment of the size proposed, including clear plans to develop capacity for potential future Phase II operations

Clear measures of success, both for Phase I operations and development of capability for a potential Phase II

Plan to evaluate with respect to those measures by gathering quantitative and qualitative data

Clear plan for thoughtful, ongoing assessment of all Institute activities

Assessment used to inform and improve both daily Institute operations and long range planning, aiming toward a successful Phase II Institute
CONTACT INFO

- Nandini Kannan, Program Director, Division of Mathematical Sciences, telephone: (703) 292-8104, email: tripods@nsf.gov
- Tracy Kimbrel, Program Director, Division of Computing and Communication Foundations, telephone: (703) 292-8910, email: tripods@nsf.gov
- Jack Snoeyink, Program Director, Division of Computing and Communication Foundations, telephone: (703) 292-8910, email: tripods@nsf.gov
- Christopher W. Stark, Program Director, Division of Mathematical Sciences, telephone: (703) 292-4869, email: tripods@nsf.gov
- Chaitanya Baru, Senior Advisor for Data Science, Directorate for Computer and Information Science and Engineering, telephone: (703) 292-4541, email: tripods@nsf.gov
QUESTIONS?
Q. I work in field A, and my co-PIs work in fields B and C. Do we meet the requirement for “significant and integral participation" by all three of the statistics, mathematics, and theoretical computer science communities?

A. It is up to the proposers to make the argument that the PIs provide expertise necessary to meet the program's goals.
Q. Is it necessary to engage in all of the “center-like” activities listed in the solicitation?

A. It is expected that the anticipated Phase II Institutes will engage in most or all of these activities. The smaller Phase I projects are expected to concentrate on some aspects of both research and education, while not necessarily addressing all the aspects listed in the Phase I solicitation.