HARNESSING THE DATA REVOLUTION (HDR): TRANSDISCIPLINARY RESEARCH IN PRINCIPLES OF DATA SCIENCE (TRIPODS)

WEBINAR

March 18, 2019

https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505347

AGENDA (HDR TRIPODS PHASE I)

- Program Directors Present
  Nandini Kannan (DMS), Tracy Kimbrel (CCF), Akbar Sayeed & Tony Kuh (ECCS)
- Overview of the Program, including what's new
- Eligibility Criteria
- Preparation Guidelines
- TRIPODS Specific Review Criteria
- Q&A

OVERVIEW

- Collaboration between:
  - **Computing and Communication Foundations** (CCF) Division in the Directorate for Computer & Information Science & Engineering (CISE),
  - **Division of Mathematical Sciences** (DMS) in the Directorate for Mathematical and Physical Sciences (MPS), and
  - **Electrical, Communications and Cyber Systems** (ECCS) Division in the Directorate for Engineering (ENG)

- Focuses on the **theoretical and foundational principles of data science** drawing on a number disciplines, including: *mathematics, computer science, statistics, information theory, signal processing, control theory, and network science*

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"

PROGRAM GOAL

HDR TRIPODS (Transdisciplinary Research In Principles Of Data Science) aims to bring together researchers from **electrical engineering, mathematics, statistics, and computer science** communities to develop the **theoretical and foundational principles of data science** through institutes for **integrated research and training activities**.

TRIPODS PHASE I

- Supports the development of small collaborative Institutes that will bring together the four disciplines.

- Proposals must address research and training in the foundational principles of data science, and demonstrate significant involvement of **at least three** of the four communities.

- Approximately 14-20 awards expected

TRIPODS PHASE I

- Teams will develop capacity and demonstrate the ability to scale activities for full Institute operations by operating as smaller Institutes

- Traditional center-like activities are anticipated
  - workshops
  - training of students & postdocs
  - workforce development
  - community building & engagement

- Novel approaches encouraged

TRIPODS PHASE I → II

- Encourage PIs to leverage existing NSF investments as appropriate through collaborations or partnerships
  - Big Data Regional Innovation Hubs
  - Engineering Research Centers
  - Mathematical Sciences Research Institutes
  - Cyberinfrastructure 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
- Tools from a variety of disciplines can be leveraged, including mathematics, statistics, computer science, information theory, signal processing, control theory, and network science.

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
- Data representation including dimension reduction and compression
- Algorithmic fairness, transparency, and interpretability
- Network influence and contagion processes
- Causal inference and artificial intelligence
- Real-time sensing and decision making for dynamic and streaming data
- Broadening machine learning with tools from control systems, information theory, and signal processing
- Integration of model-driven and data-driven approaches

ELIGIBILITY CRITERIA

INSTITUTIONS

- Any institution that received an award under the previous TRIPODS Phase I solicitation NSF 16-615 **is not eligible to submit a single-institution proposal**.

- These institutions may submit only as either:
  a) lead or non-lead collaborative partners on a multi-institution collaborative proposal, or
  b) sub-awardees

WHO MAY SERVE AS PI

- No PI, co-PI, or Senior Personnel for an award under the previous TRIPODS Phase I solicitation NSF 16-615 may serve as a PI, co-PI, or Senior Personnel.

LIMIT ON NUMBER OF PROPOSAL PER PI

- An individual may serve as PI or co-PI on at most one project team but may serve as other Senior Personnel on any number of teams.
- In the event that an individual exceeds the PI or co-PI limit, proposals will be accepted based on earliest date and time of proposal submission, i.e., the first proposal will be accepted, and the remainder will be returned without review.
- No exceptions will be made.

PROPOSAL PREPARATION GUIDELINES

PROPOSAL PREPARATION: DEADLINES AND TITLE

Required Elements (in addition to the requirements listed in the NSF PAPPG)

- **Letter of intent required**
  - Due date: March 25, 2019
- **Full proposals**:
  - Submission window: April 24-May 8, 2019
- **Proposal Title**: Proposal titles should begin with "HDR TRIPODS" followed by a colon, then the title of the project, i.e., **"HDR TRIPODS: Title".**

PROPOSAL PREPARATION: COLLABORATION & EVALUATION PLAN

- **Collaboration and Evaluation Plan (Required):**
  - Describe expertise in three of the four 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

The Collaboration and Evaluation Plan must be uploaded as a separate Supplementary Document (limited to 5 pages). Proposals without this document will be returned without review.

PROPOSAL PREPARATION: PROJECT DESCRIPTION

- 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**

TRIPODS SPECIFIC REVIEW CRITERIA

SOLICITATION SPECIFIC REVIEW CRITERIA

- **Well-integrated research and training program** focused on the theoretical foundations of data science and **fostering collaboration and interaction among the four communities of TRIPODS** – electrical engineering, mathematics, statistics, and theoretical computer science

- Address the "**Broad themes of the program**"

- Strategies for **workforce development, including novel educational and training activities**

- **Transdisciplinarity/Synergy**:
  - Bringing together theories, techniques and principles from electrical engineering, mathematics, statistics, and theoretical computer science
  - Synergy between the groups

SOLICITATION SPECIFIC REVIEW CRITERIA

- **Vision**: Ability to identify and articulate a vision for the foundational principles 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 expertise** represents at least three of the four communities

- Well-developed plan for **communication and interaction with the domain areas 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**

SPECIFIC REVIEW CRITERIA – COLLABORATION & EVALUATION PLAN

- Clear **measures of success**, both for Phase I operations and **development of capability for a potential Phase II**
- Plans to evaluate with respect to those measures by **gathering quantitative and qualitative data**
- Clear plan for **thoughtful, ongoing assessment of all Institute activities**
- The use of assessment to inform and improve both daily Institute operations and long range planning, aiming toward a successful Phase II Institute

CONTACT INFORMATION

- Nandini Kannan, Program Director, Division of Mathematical Sciences, telephone: (703) 292-8104, email: nakannan@nsf.gov
- Tracy Kimbrel, Program Director, Division of Computing and Communication Foundations, telephone: (703) 292-7924, email: tkimbrel@nsf.gov
- Anthony Kuh, Program Director, Division of Electrical, Communications, and Cyber Systems, telephone: (703) 292-2210, email: akuh@nsf.gov
- Phillip A Regalia, Program Director, Division of Computing and Communication Foundations, telephone: (703) 292-2981, email: pregalia@nsf.gov
- Akbar Sayeed, Division of Electrical, Communications and Cyber Systems, telephone: 703-292-4753, email: asayeed@nsf.gov
- Rahul Shah, Program Director, Division of Computing and Communication Foundations, telephone: (703) 292-2709, email: rshah@nsf.gov
- Christopher W. Stark, Program Director, Division of Mathematical Sciences, telephone: (703) 292-4869, email: cstark@nsf.gov

QUESTIONS?

Q&A 1

- 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 at least three of the electrical engineering mathematics, statistics, 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&A 2

- 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.