

Side 1 –

Good afternoon, and welcome to the webinar for the Transdisciplinary Research in Principles of Data Science, or TRIPODS program. I'm Tracy Kimbrel, Program Director in the Computing and Communications Foundations division.

Slide 2 - Agenda

Here's the agenda for today's webinar. The CISE and MPS Assistant Directors, Jim Kurose and Fleming Crim, will give opening remarks. The Division Directors for CCF and DMS, Rao Kosaraju and Michael Vogelius, are also on hand— we are all very excited about this program. Next, Nandini Kannan, Program Director in the Division of Mathematical Sciences, will give an overview of the program. Then I'll go over some specific things you'll need to pay attention to to submit a proposal, including both administrative details and the criteria that will be used to evaluate the proposals. Finally, Jack Snoeyink, Chris Stark, and Chaitan Baru will also be on hand to take questions.

(Jim and Fleming speak)

Slide 3 - Overview

Good afternoon, I'm Nandini Kannan, program director in the Division of Mathematical Sciences, and I'll provide an overview of the TRIPODS program.

TRIPODS is a joint program with equal participation between the Computing and Communication Foundations Division and the Division of Mathematical Sciences. The program supports one of the ten big areas called out by the Director of NSF and her team, "Harnessing Data for 21st Century Science and Engineering." TRIPODS is one of many NSF programs that support this area; we will touch briefly on some of the others later. TRIPODS focuses in particular on the algorithmic,

mathematical, and statistical foundations of data science.

Slide 4 – Need for TRIPODS

In May of this year, CCF and DMS sponsored a workshop on this topic, called “Theoretical Foundations of Data Science: Algorithmic, Mathematical, and Statistical,” which brought together researchers from the three areas to consider what the communities should be doing and how NSF might support it. The workshop participants produced a report, and you can find a link to it in the TRIPODS solicitation, which we hope you have already read carefully. The key conclusions of the workshop report are that theoretical foundations are important throughout the data lifecycle and are critical for data science, and that this is an interdisciplinary topic that will need multiple communities to collaborate across traditional boundaries.

Slide 5 – Program Goal

In response to this need, CCF and DMS developed the TRIPODS program. The aim is to bring together and integrally involve all three of the statistics, mathematics, and theoretical computer science communities. You may notice that we list these three in different orders throughout the solicitation and this presentation,. This is purposeful. No one of these should be seen as having any more or less important role than any of the others, and as we repeat here and in the solicitation, all three must be involved in TRIPODS.

We envision center-like activities integrating research, education, and training, at what we will call “institutes,” because “center” has a specific connotation at NSF .

Slide 6 – TRIPODS Phase 1

The current solicitation is for what we call phase one. In phase one, we intend to support eight to ten smaller institutes, at a level of funding of up to five hundred thousand dollars per year for three years, dependent on availability of funds . As mentioned, we expect all three communities to be involved, and proposals should address both research and training aspects of an institute for principles of data science.

Slide 7 – TRIPODS Phase 1

The idea is for phase one to allow teams to develop capacity for larger-scale operations and demonstrate their ability in anticipation of phase two. We expect to see some of the usual approaches and activities associated with research and training centers, and we encourage some out-of-the-box thinking as well.

Slide 8 – TRIPODS Phase I to II

We also encourage TRIPODS institutes to engage and take advantage of existing NSF investments such as Big Data Hubs, Math Institutes, and Software Institutes supported by the Software Infrastructure for Sustained Innovation program.

Toward the end of the three years of phase one, we anticipate a second competition we will call phase two (naturally enough), subject to availability of funds in years 2020 and beyond. For phase two, the idea is to select a smaller number of teams from among the phase one institutes for larger awards over a longer period. The precise rule is that at least one PI or co-PI of a phase two proposal has to have been a phase one PI or co-PI. We will encourage a lot of interaction among the phase one institutes, and in fact will encourage them to consider joining forces to form

larger teams for the phase two competition.

Slide 9 – Broad Themes of the Program

Now I'll describe a little bit of detail on what we are expecting in terms of the activities we expect TRIPODS institutes to engage in. We want to see the three communities come together to overcome barriers to collaboration such as using different jargon and models for overlapping ideas and methods. Applicability of foundational research is critical, so connections to domain scientists and industry will be key. Foundational aspects of data science education are an important consideration, but these should be developed to include not just abstract computational issues but also experimental validation, ethical behavior, and interdisciplinary communication. Finally, some of the most important issues revolve not around how to analyze a clean and complete data set, but the entire pipeline from continuing dynamic acquisition of noisy data through decision making to data archival or retirement.

Slide 10 - Possible Research Foci

The workshop report and the solicitation provide a number of possible research areas for the three communities to focus on as they come together to address foundational aspects of data science. These are listed on the slide, and I won't read them out loud. I want to emphasize that these are just examples suggested by the participants of a two-day workshop. These are in no way exhaustive, and should just serve as examples to get your imagination started. And we emphasize again that all three communities – mathematics, theoretical computer science, and statistics – must be integrally involved and engaged.

Slide 11 – Proposal Preparation & Review Criteria

(Tracy) Thanks, Nandini. Now I'll give some details on what's required in TRIPODS proposals and how they will be evaluated.

Slide 12 – Proposal Preparation Guidelines

Of course, all proposals are subject to the NSF Proposal and Award Policies and Procedures Guide (the PAPPG; you may hear this referred to as the PAPP Guide). For this program, we're requiring letters of intent to give us some advance warning of the proposals to come a bit later. What we're really after here is a brief summary of what you plan to propose, and, importantly, lists of involved personnel, so we can start planning the review process. These letters are due January 19th.

Full proposals are due March 15th. In addition to the usual fifteen page Project Description, we're asking for a Collaboration and Evaluation Plan of up to five pages. This document should describe the co-PIs and their expertise, along with their plans for collaborating to achieve their objectives and meet the goals of the program. It should also give clear measures of success for the project, along with plans for evaluating the project against those measures.

Slide 13 – Proposal Preparation Guidelines

The main section of the proposal (the Project Description) should include (among other things) a timeline for the project and should demonstrate how the team will gear up for full-scale operations in phase two.

We have a new way for you to submit your collaborators and other affiliations that I want to call your attention to. You will need to enter this information into a spreadsheet, and we will provide a template for it. Please see the solicitation for details.

Slide 14 – Solicitation Specific Review Criteria

In addition to the standard review criteria Intellectual Merit and Broader Impacts used to evaluate all NSF proposals, we've defined several special criteria for TRIPODS. Reviewers will be asked to familiarize themselves with these and consider all of them when evaluating proposals. These include the quality of the proposed plan for integrated research and training activities, the project's plans to foster collaboration among the three disciplines, the broad themes of the program mentioned previously (and which are detailed in the solicitation that you should read carefully), and the workforce development and educational aspects of the project.

Slide 15 – Solicitation Specific Review Criteria

Review criteria also include the synergy and transdisciplinarity across the three disciplines, the project's vision for the emerging field of data science, whether the PIs' backgrounds are appropriate and complementary and whether their roles are clear, and the plans to interact with domain sciences and industry to ensure relevance of the foundational development planned for the project.

Slide 16 – Solicitation Specific Review Criteria

Proposals will also be evaluated on their plans to develop capacity for full-size operations in phase two, and whether they include clear measures of success as mentioned previously, for phase one operations as well as plans to gear up for phase two, along with plans to evaluate success against those measures, thoroughly assess all aspects of the institute's activities, and feed back those evaluations to ensure the long term success of the project.

Slide 17 – Contact Info

Here is contact information for the TRiPODS program director team. This is also available in the solicitation, which, of course,

you should read carefully. The slides and transcript for today's presentation will be posted shortly on the event page where you registered for the webinar, so don't worry about trying to copy all these before I advance to the next slide.

Slide 18 - Questions

Thanks for your attention. Now, Chaitan, Jack and Chris will join us to take questions from the audience. (Operator gives instructions. Chaitan, Jack, and Chris introduce themselves.)

While we're waiting for questions to come in, we'll answer a couple of questions that we've prepared.

Slide 19 – Q&A 1

Q: I work in field A, and my co-PIs work in fields B and C (or I am in the Dept. of D, and my co-PIs are in the Depts. of E and F; or I have a PhD in G, and my co-PIs' degrees are in H and I). 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. This should be detailed in the Collaboration and Evaluation Plan. Please read the solicitation carefully for more information on this and other aspects of submitting a competitive proposal.

Slide 20 – Q&A 2

Q: Is it necessary to engage in all of the "center-like" activities listed in the solicitation?

A: The envisioned Phase II TRIPODS Institutes will conduct comprehensive research and education programs in foundations of

data science. 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. Phase I projects are intended to develop and demonstrate the capacity to operate a Phase II Institute. This may involve significant effort in a subset of the activities, along with more limited effort, or a credible plan to develop capacity, in other listed activities.