Slide 1: Harnessing the Data Revolution (HDR): Data Science Corps (DSC)
<Aidong>
Good Morning
I am delighted to welcome all of you to this webinar on Harnessing the Data Revolution (HDR): Data Science Corps (DSC), solicitation NSF 19-518.

We have received many questions since the solicitation is posted. I hope you find this webinar informative.

Slide 2: Agenda
<Aidong>

Here is an outline of today's presentation. We will start with the HDR program introduction by the leaders of HDR steering committee, Dr. Manish Parashar.
We will then give an overview of the specific HDR Data Science Corps program.

The Cognizant Program Directors for HDR DSC include Stephanie August from the Division of Undergraduate Education in the Directorate of Education and Human Resources Directorate, Cheryl Eavey from the division of Social and Economic Sciences of the Directorate of Social, Behavioral, and Economic Sciences, Nandini Kannan from the Division of Mathematical Sciences in the Directorate of Mathematics and Physical Sciences, and myself, Aidong Zhang, from the division of information and intelligent Systems of the Directorate of Computer and Information Science and Engineering. In this webcast, we will provide a brief overview of the data science corps program, and describe some of the most important things you need to know about submitting a proposal.

We will then take questions from the audience.

I will now turn it over to Manish who is co-leading the HDR steering committee.

Slide 3:
<Manish>
Good Morning

On behalf of the HDR team, it is my pleasure to welcome you to this webinar.
As you may know, in 2016, the National Science Foundation (NSF) unveiled a set of "Big Ideas," 10 bold, long-term research and process ideas that identify areas for future investment at the frontiers of science and engineering. The Big Ideas represent unique opportunities to position our Nation at the cutting edge of global science and engineering leadership by bringing together diverse disciplinary perspectives to support convergence research.
HDR is one of the 10 Big Ideas.

Slide 4: Harnessing the Data Revolution Vision
< Manish>

HDR is a national-scale activity to enable new modes of data-driven discovery that will allow new fundamental questions to be asked and answered at the frontiers of science and engineering. Through this NSF-wide activity, HDR will generate new knowledge and understanding, and accelerate discovery and innovation.
The HDR vision is realized through an interrelated set of efforts focused on:

- The foundations of data science;
- Algorithms and systems for data science;
- Data-intensive science and engineering;
- Data cyberinfrastructure; and
- Education and workforce development.

Slide 5: Harnessing the Data Revolution (HDR): Solicitations
< Manish>
The Data Science Corps solicitation, which is topic of this webinar, focuses on education and workforce development.
I would like to briefly mention a second solicitation that focuses on the conceptualization of HDR Institutes for Data-Intensive Research in Science and Engineering.

The HDR Institutes activity seeks to create an integrated fabric of interrelated institutes that can accelerate discovery and innovation in multiple areas of data-intensive science and engineering by harnessing diverse data sources and developing and applying new methodologies, technologies, and cyberinfrastructure for data management and analysis.

I will now turn it back to Aidong.

Slide 6: Data Science Corps Program
<Aidong>
Next we talk about this specific solicitation opportunity (NSF 19-518). As Manish just mentioned, the Data Science Corps (DSC) program is the education and workforce development component of the NSF Harnessing the Data Revolution (HDR) ecosystem.

To train the next generation of students in Data Science, HDR DSC provides support for integration of real-world data science projects into classroom instruction.

In particular, undergraduate students receive stipends ($5000 per student) to participate in data science projects.

Slide 7: Data Science Corps Goals
<Aidong>
I would like to emphasize the two specific goals of this solicitation are:

- Build capacity for harnessing the data revolution at the local, state, national, and international levels to help unleash the power of data in the service of science and society.

- Provide practical experiences, teach new skills, and offer teaching opportunities, in a variety of settings, to data scientists and data science students.

The set of awards made by this program are expected to reflect geographic diversity and represent the wide variety of regions across the country. Stephanie will talk about the team requirement in the next slide.

Slide 8: Data Science Corps Team Requirements
<Stephanie>
Let's turn now toward the requirements for the composition of each Data Science Corps Team.

Collaboration is a key component of Data Science Corps projects on many levels. Each project team is expected to engage two or more organizations, with one assuming the role of coordinator and the other assuming the role of implementor. The coordinating institution will coordinate, monitor, and evaluate all project activities. It will also provide faculty professional development and mentoring.

One or more *implementation* organizations should be affiliated with each *coordinating* organization. The implementing institution will implement the Data Science Corps in one or more data science courses and provide stipends to support the undergraduate participants. The coordinating institution and the implementation institutions are expected to work together on the introduction of Data Science Corps activities into existing courses, identifying project organizations, recruiting suitable projects, and developing a common set of criteria for awarding Data Science Corp stipends to students.

One organization can assume both roles, but keep in mind the expectation that each project will involve multiple organizations. We encourage you to refer to the solicitation for a detailed description of the coordinating and implementation roles.

Slide 9: Submission Limits
<Stephanie>
Now let's review limits on submissions.

While project teams are expected to engage multiple organizations, each institution is allowed to submit only one proposal. On collaborative projects, both the lead proposal and each collaborative proposal count as a submitted proposal. Subawards, on the other hand, do not count toward the limit.

Likewise, each individual may participate in only one project as a PI, Co-PI, or senior personnel. Both lead and collaborative proposals count. Being a subawardee on a project does not count toward the limit.

Note that a subawardee organization does not submit a proposal, and personnel on a subaward are not the PI, Co-PI, or Senior Personnel on the proposal unless the individuals happen to be also named as a Co-PI on the proposal's Cover Sheet. An organization could appear as a subawardee in any number of proposals, and could also submit a proposal itself.

If multiple submissions are received from an institution or an individual, the first submission received from an institution or individual will be retained. Subsequent submissions will be returned without review.

Slide 10: PROPOSAL PREPARATION GUIDELINES
<Aidong>

Next we talk about the proposal preparation guidelines.

First, full proposals are required to be submitted between January 28, 2019 and February 4, 2019, due by 5pm submitter's local time on February 4, 2019.

Proposal preparation should follow the instructions described in the NSF 18-1 PAPPG (Proposal & Award Policies & Procedure Guide), such as 15 page limit and other format requirements.

Next we talk about program specific requirements in proposal preparation.

Slide 11: Project description
<Aidong>

Proposal should focus on what and how the supported undergraduate students will participate in any data science related projects, such as real-world projects in the local community, industry, government, academia, non-profit, and other sectors, to gain hands-on experiences.

For example, proposal may incorporate it as part of a data science curriculum, such as capstone projects, summer internships, coop programs, study abroad programs, or similar experiences.

The description of the specific roles in the team mentioned earlier should be detailed here.

The assessment of student learning and program evaluation is a critical component. Stephanie will talk about it in the next slide.

Slide 12: Assessment of student learning and program evaluation are required
<Stephanie>
As we mentioned earlier, the coordinating institution is responsible for the assessment of student learning, and evaluation of the program overall. We can think of assessment as a chef tasting a soup as she is preparing it, adjusting ingredients and seasonings as needed while the soup is cooking to achieve the desired result. Evaluation, on the other hand, occurs when the customer tastes the soup and provides feedback on the result.

Assessment should aim to address the impact of the hands-on experiences on student gains in content knowledge related to data science, as well as gains in the translational skills needed to enable effective application of data science content knowledge. These translational or "super skills" include teamwork, entrepreneurship, critical thinking, communications, collaboration, creativity, and, perhaps most importantly, ethics. Assessment is also expected to examine student persistence as a STEM major and persistence in a STEM-related career path. Note that this does require a longitudinal study into the value

of data science, and the Data Science Corps, as a means of increasing the ranks of the national STEM workforce.

Project evaluation efforts should address issues related to the overall impact of the *Data Science Corps.* Let's look at few examples of relevant evaluation questions:

- Did the award reach its goals? How do we know this, or why was it not successful in reaching its goals?
- What are the characteristics of an effective *Data Science Corps* award, including specific projects or project organizations?
- Has the award been effective at developing models for communication and engagement among disciplines at the student level? What about communication and engagement among faculty, and broadly within the workforce?
- Are certain types of *Data Science Corps* projects more effective for certain students, or for certain organizations, and/or for certain communities? And finally,
- What value do specific projects produce for the communities they serve?

You are encouraged to refer to *The 2010 User-Friendly Handbook of Project Evaluation*, which is cited in the solicitation. The *Handbook* discusses quantitative and qualitative evaluation methods and is a valuable resource for understanding the value of evaluation and how to design and carry out an evaluation.

Slide 13: Budget
<Stephanie>
What does a budget for a Data Science Corps project look like?

A total of $10 million has been allocated for this program. We expect to make eight to eleven awards ranging from $1 to $1.2 million for projects with duration of up to three years. Awards are expected to support about 25-30 students per year across all participating organizations, at approximately $5,000 per student, budgeted under Participant Support. Students must be citizens or permanent residents of the U.S or its territories. The overall project budget in year three of an award is expected to be less than in years 1 and 2, in anticipation of transitioning the award to other source(s) of funds.

The budget for the coordinating organization may include support for the principal investigator, full or partial support for a project coordinator, project communications, program evaluation, appropriate travel expenses, and usual administration and other expenses.

The budget for each implementation organization should include student stipends, and may also include support for the award PI or co-PI and other staff support. If students are required to travel to project locations, additional travel support may be included, with proper justification.

Slide 14: PROPOSAL REVIEW General review criteria
<Aidong>
Next we talk about general proposal review criteria.

Slide 15: NSF REVIEW CRITERIA
<Aidong>

As for all proposals received by NSF, DSC reviewers and panelists will be asked to consider the intellectual merit and broader impact for each proposal for their review, panel discussions, and panel summaries.

To reiterate these criteria:

Intellectual Merit: Encompasses the potential to advance knowledge, in this case, to build data science workforce.

Broader Impacts: Encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes. We talk about this in detail in the next slide.

Slide 16: NSF Broader Impacts
<Aidong>

The following aspects should be discussed/addressed**:**

- participation of women, persons with disabilities, and members of **underrepresented groups**
- science, technology, engineering, and mathematics **education**
- increased **public scientific literacy and engagement** with science and technology
- development of a **diverse, globally competitive workforce**
- increased **partnerships between academia and industry**
- improved national **health, security and economic competitiveness**
- enhanced **infrastructure for research and education**

Next Stephanie will talk about the solicitation specific review criteria.

Slide 17: SOLICITATION-SPECIFIC REVIEW CRITERIA
<Stephanie>
There are four additional solicitation-specific review criteria to consider.

Slide 18: Diverse participation linked to courses and training
<Stephanie>
Data Science Corps projects are expected to represent diversity and be well integrated with their institutions.

First, proposals will be evaluated on their organizational diversity, that is, whether the proposal reflects diversity among participating organizations. This might be accomplished by reaching across many types of institutions of higher education, including research universities, two- and four-year colleges, and minority serving institutions.

Second, proposals will be evaluated on the clarity of their linkages to existing or new undergraduate courses, as well as internship, traineeship, study aboard, and/or other relevant programs at the institutions.

Slide 19: Real-world experience and community connections
<Stephanie>
Data Science Corps projects are expected to reflect real-world experience and community connection.

Third, proposals will be evaluated on how well the proposing teams demonstrate a track record and prior experience with implementing data science projects in academia, industry, government, non-profit, and/or other relevant sectors. Teams are also expected to demonstrate the ability to clearly identify specific projects in those settings.

Lastly, proposals will be evaluated on their connection to stakeholder communities. The proposing team should demonstrate linkages with relevant stakeholders in industry; local, state, and/or federal governments; non-profits; international organizations; and/or other sectors.

Slide 20: Solicitation-Specific Supplementary Documents
<Stephanie>
HDR DSC proposals are expected to include two solicitation-specific supplementary documents.
A schedule of activities, limited to two pages, should describe the schedule of meetings among collaborators. It should also include a tentative schedule for quarterly or more frequent teleconference meetings and annual or more frequent in-person meetings among all Data Science Corps awardees, including the annual Data Science Corps meeting in the Washington, DC, area. All tentative schedules can be made firm after all projects have been awarded and announced.

Proposals are expected to include a letter of collaboration from each participating organization named in the proposal, other than the submitting lead, non-lead, and/or subawardee institutions. Letters should explicitly state the nature of the collaboration, appear on the organization's letterhead, and be signed by the appropriate organizational representative. These letters must not otherwise deviate from the restrictions and requirements specified in the NSF Proposal and Award Policies and Procedures Guide, which you can find by searching for NSF 18-1 on the web.

Slide 21: Questions?
<Aidong>

We have now completed the formal portion of the presentation. Before opening the telephone lines to questions from the audience, we would like to address a few of the questions we have already received.

Slide 22: Q&A 1
<Aidong>

Q1: Given the institutional proposal limit of one proposal submission per organization, would a non-lead collaborative proposal submission from an organization (i.e., not as a subawardee) count against that non-lead organization's institutional proposal limit?

Answer: yes

Slide 23: Q&A 2
<Aidong>

Q2: The solicitation calls for a "coordination organization" and one or more "implementation organization(s)". Must organizations be different universities or institutions, different colleges, different departments?

Answer: **No**. But the solicitation also says, “participation by more than one organization is expected in each *Data Science Corps* award.” So if much of the activity in the project is concentrated at one university, that will work against the project in the review process.

Slide 24: Q&A 3
<Nandini/Aidong>

Q. Do all the undergraduate students/participants need to be data science majors?

A. **No**. Students enrolled in any STEM degree program are eligible to participate

Thanks for your attention. Now we will take questions from the audience. (Operator gives instructions.)

Slide 25: Thank you!
Thank you for your interest in the Data Science Corps. We encourage you to read the solicitation carefully, and to make a checklist of the components and requirements before writing your proposals. You are welcome to email any questions that are not answered in the solicitation or on the program web page. For best results, we recommend that you direct your email to a single program officer. We look forward to receiving your proposals.