Slide 1:

Welcome everyone to the webinar today. My name is Fen Zhao, Program Coordinator at NSF in the Directorate for Computer and Information Sciences and Engineering, Office of the Assistant Director. Today I'll be giving you some background on the BD Hubs and Spokes program, and the 2017 BD Spokes solicitation

Slide 2:

Before we start, I want to make sure to acknowledge the whole team here at NSF that has been working on the BD Hubs program here in the NSF CISE Directorate.

Slide 3:

The agenda for our webinar today is three fold. I'll run start by running through some the historical timeline and background surrounding the BD Hubs and BD Spokes program, setting some context on where this program sits within broader NSF and CISE efforts, as well as the interaction between the established BD Hubs and the new BD Spokes to be funded.

Second, I'll try to explain some of the overall strategy and motivation behind this BD Spokes program and the major themes and topic areas of interest to NSF. The thrust behind the BD Spokes solicitation is accelerating Big Data applications, which topically has a very large scope. The thematic areas articulated in the solicitation help guide the proposer to understand the dimension where NSF feels innovation in Big Data could make the most impact.

Finally, I will discuss some of the logistical details related to the proposal and review process for the 2017 BD Spokes solicitation.

Slide 4:

The BD Hubs and Spokes comprise just one component of a broader Big Data portfolio of programs that NSF funds. Here at NSF we have a four-part framework of how we think about Big Data. One is the foundational research that's needed, which at NSF is funded primarily by the Critical Techniques and Technologies for Big Data program (often just called BIGDATA, all caps). We also think about supporting infrastructure design and development, which is primarily supported by the Data Infrastructure Building Blocks or DIBBS program. We also fund education in Big Data through multiple programs, including, in the past the National Research Traineeship. Finally, on the partnerships and engagement quadrant of the portfolio, we have the BD Hubs and Spokes program.

So as you listen through this webinar, you may find that the unique nature of the BD Spokes isn't a good fit for the project you have in mind. I encourage you to take a look at the other programs we mention today.

Slide 5: The BD Spokes solicitations are part of the Big Data Regional Innovation Hubs Program (also known as BD Hubs). The BD Hubs program was launched in March of 2015 with a call to develop 4 Big Data hubs around the country. NSF held 4 regional charrettes gathering stakeholders from each region, who collaborated to submit a single BD Hub proposal to NSF in June 2015. In September 2015, NSF announced four awards to establish four regional hubs for data science innovation. I will have more information about the BD Hubs in later slides.

The BD Spokes solicitations kick off the second phase of the BD Hubs program. The 1st solicitation was released in Nov 2015, and $12 million was awarded. In March 2017, we released the 2nd BD Spokes solicitation -- the basis of what this webinar.

Slide 8:

This slide shows the awards made by the 1st BD Spokes solicitation. We funded a wide variety of projects that align with the regional priorities, for example: transportation, manufacturing, health, or environmental, agriculture, and education. There were 11 Spoke grants and 10 Planning grants awarded. More information about individual Spoke can be found on the corresponding BD Hub website and on the NSF Awards Database.

Slide 9:

The Hubs consortia are coordinated by top data scientists at Columbia University (Northeast Hub), Georgia Institute of Technology and the University of North Carolina (South Hub), the University of Illinois at Urbana-Champaign (Midwest Hub) and the University of California, San Diego, the University of California, Berkeley, and the University of Washington (West Hub).

Covering all 50 states, they include commitments from universities and cities to foundations and Fortune 500 corporations---with the ability to expand further over time. The organizational structure for each BD Hub varies by region. Some are organized into standing sub committees and task forces, others into flexible working groups. You see on this map not only the Hub coordinators, but also the organizations leading the subcommittees and sub divisions of each Hub.

Slide 10:

Now, let's turn to the specifics of the 2017 BD Spokes solicitation...

Slide 11:

BD Spokes proposals must articulate a clear focus within a specific Big Data topic or application area, while highlighting their Big Data Innovation theme. All BD Spokes must have clearly defined mission statements with goals and corresponding metrics of success. Some templates illustrating the specificity and level of detail for missions include:

- Use a specific set of analytical tools to improve the lead time for predictions of certain critical regional indicators by a given percentage.

- Given a specific set of high value data sets that were previously siloed and, therefore, usable only within a single research group or institution, make them available to a broader set of groups, or to the public at large, along with appropriate privacy and access control mechanisms.

- Adapt specified Big Data technologies to automate previously tedious and manual data collection and curation processes for specific types of data in a given field of science.

- For a specific genre of data, introduce new types of (automated) analytics—which were previously tedious to perform and manual in nature—that can be performed with minimal

BD Spokes can initiate many different kinds of activities in support of their mission goals. The BD Spokes role is meant to convene stakeholders to augment and spawn new research efforts as opposed to directly carrying out traditional research. Potential activities for BD Spokes include, for example:

- Accelerating the creation and development of Big Data solutions relevant to its mission by convening stakeholders across sectors (e.g., academic, industry, non-profits, etc.) to partner in results-driven programs and projects;

- Driving successful pilot programs by acting as a matchmaker between the various academic, industry, and community stakeholders;

- Engaging stakeholders across the region—including solution providers and end users—to enable dialogue, share best practices, and/or set standards for data access, data formats, metadata, etc.; and

- Connecting critical data resources to stakeholders that can best utilize them to fulfill the BD Spoke mission.

Note that BD Spokes funding from this solicitation is not intended to primarily support research activities. Rather, the goal of the program is to enhance and amplify collaborative efforts focused on achieving specific mission-based goals. For specific questions about appropriate activities for BD spokes proposals, please reach out to me.

Slide 12:

Similar to the 1st BD Spokes solicitation, the activities of a 2017 BD Spoke should also address one or more of the following Big Data Innovation themes:

- *Accelerating progress towards societal grand challenges relevant to regional and national priority areas.*
- *Helping to automate the Big Data lifecycle.*
- *Enabling access to and increasing the use of important and valuable available data assets, also including international data sets, where relevant.*

Note that this list is not MECE, mutually exclusive or collectively exhaustive. These are simply a few critical dimension we find to be some priorities in accelerating innovation in Big Data. Any proposal may touch on more than one of these themes. There may also be Big Data proposals that make an important impact on data innovation in a way that this framework does not capture, however, for this competition, those projects would not be a good fit.

Slide 13:

The first theme is "accelerating progress towards societal grand challenges relevant to regional and national priority areas". Due to the pervasiveness of Big Data in virtually all national priority areas, the BD Spokes have the opportunity to bring rapid change in application areas, by facilitating the creation of interdisciplinary and multidisciplinary data-intensive teams. Such areas broadly include healthcare, climate change, urban sciences and many more. These grand challenges may have widespread global implications, but be of specific interest to the region. For example, the South Hub has articulated an interest in their region on coastal hazards, in particular of the gulf coast. While projects on coastal

Slide 14:

The second theme is "helping to automate the Big Data lifecycle." Managing the end‑to‑end lifecycle of Big Data assets can be a tedious and manual task. Steps in the data lifecycle include: ingestion, validation, curation, quality assessment, anonymization, publication, active data management, and analysis (including information extraction, visualization, and annotation). Automated (or, semi‑automated) techniques are needed in order to keep up with the rapid data rates, large volumes, and immense heterogeneity of Big Data. Automation may also aid the reproducibility of data processing and analysis workflows. The data challenges and lessons learned by a BD Spoke on such automation efforts are expected to be shared with the BD Spoke's stakeholders as well as more broadly across the network of BD Hubs and Spokes.

Slide 15:

The third theme is "enabling access to and increasing the use of important and valuable available data assets, also including international data sets, where relevant." Many valuable data sets are underutilized, and results from the analysis of such data are not shared, due to a variety of actual or perceived costs, including cost of curation, cost of data reuse, attribution and intellectual property considerations, etc. One of the desirable roles for a BD Spoke is as a catalyst for organizing and sharing datasets and related data services among a larger set of stakeholders, across disciplinary areas, within the geographic region, or across the national community. BD Spokes are expected to play an important role in supporting and promulgating open data and open source software policies within their projects— at the Hub‑level, and across Spokes—to further facilitate the sharing of data and outcomes of analyses.

Broadly, for projects addressing any of the three themes, NSF recognizes that Big Data are global due to the way they are collected and analyzed and, hence, encourages international collaborations that will enhance the capacity and capabilities of the BD Hubs and Spokes.

Slide 16:

In the solicitation that was posted as NSF document 17-546, it mentioned 2 types of awards: MEDIUM and SMALL. The MEDIUM grants range from $500K to $1M while the SMALL grant is for proposal between $100 and $500K, both are up to 3 years. We anticipate a total funding amount of $10M for the program. This would come to approximately 10 to 20 awards, though distribution would depend on quality and volume of proposal in these two categories.

The Full proposal deadline is Sept 18, 2017, 5pm proposer's local time.

NO Letter of Intent is required. However, the proposer must have communicated with their collaborating hub and obtained a Letter of Collaboration.

Slide 17:

For MEDIUM proposals, which range in budget from $500K to $1M, you must be able to articulate tangible outcomes. Here are some examples of what we meant by tangible outcomes:

- **Explicit results** from data-enabled or data-facilitated inquiry in a scientific or engineering field or other

domain-specific questions addressed and which specific researchers would be conducting the research to answer these questions

- A **prototype** or proof of concept for a technology platform, data product, data standards, or other data infrastructure

- An innovative **education or workforce development program** with a plan for evaluating the effectiveness of the program. The program could be virtual-, classroom-, or workplace-based and should engage underrepresented groups.

Slide 18:

Proposed BD Spoke projects are expected to focus on their articulated regional challenges and opportunities. In addition, NSF has a working group of program officers from a number of different directorates and cross agency programs that are interested in solicitation additional proposals from the Big Data community that addresses some of their interests and priority areas. This includes:

- *Neuroscience*: Engage questions and opportunities in neuroscience that leverage BD Hub resources, such as enabling large scale, integrative modeling, sharing of diverse data and resources, and other neuroscience and neurotechnology approaches that require very large-scale, complex, or diverse data. Connections to other NSF programs on neuroscience research (www.nsf.gov/brain/) are welcome.
- *Replicability and Reproducibility in Data Science*: Facilitate robust and reliable science by improving the replicability and reproducibility of research instruments, procedures, codes and results.
- *Data-driven Research in Chemistry:* Encourage innovative partnerships that capitalize on the data revolution (https://www.nsf.gov/pubs/2017/nsf17036/nsf17036.jsp) and utilize discovery-based science to verify scientific predictions and insights in chemistry. This area of emphasis looks for formation of new alliances to accelerate the discovery of new chemical species with predicted properties and/or new chemical reactions using approaches such as large-scale data analysis, data architectures, or machine learning. Proposal topics must be in alignment with the core research programs within the Division of Chemistry (CHE; www.nsf.gov/div/index.jsp?div=che).
- *Education:* Support innovations in software infrastructure and the use of education and learning data sets arising from both administrative data and information collected from interactive learning systems to improve learning outcomes. Projects could also propose to develop innovative education and/or workforce development and training programs that both broaden participation in Big Data research and development activities and enable a workforce for the 21st century. Workforce and training activities will be evaluated on their innovativeness and their ability to be replicated in new environments.
- *Data Intensive Research in the Social, Behavioral, and Economic Sciences*: Accelerate research infrastructure and frameworks that integrate and operate on data from multiple sources including administrative data; scientific instruments from large-scale surveys, brain research, large-scale simulations, etc.; digitally-authored media, including text, images, audio, and emails; and streaming data from weblogs, videos, and financial/commercial transactions.
- *Data Analytics for Security*: Better analytics and detection of security- and privacy-related patterns, anomalies, trends and changes in BD Spoke applications and/or regional data exchanges. Development of statistical, computational and/or interdisciplinary methods for

analytics, mining, and visualization of structured or unstructured BD Spoke data from disparate sources.

Slide 19:

The review process for the solicitation will be by the standard NSF merit review process through panel or ad-hoc reviews. As this is a unique solicitation, reviewers will also address the degree to which the proposed BD Spoke:

- Needs the support and infrastructure of the BD Hub network to carry out its mission;

- Effectively coordinates with its corresponding regional BD Hub and augments the capacity of the existing National Network of BD Hubs and Spokes;

- Will contribute to the education and training plan of the regional BD Hub, including its efforts to educate and train not only the Big Data workforce but also related external groups such as end users, students, or managers; and

- For MEDIUM proposals: Has a feasible set of outcomes that can be delivered given the activities planned.

To best help a proposal address these criteria, there are a series of mandated sections in the proposal that are outline in the solicitation. I won't go into detail for them on screen because they are different small and medium proposals, but please take a look.

Slide 20:

Proposals can be submitted by a wide number of organizations such as universities, colleges, non-profits, state and local governments. NSF welcomes collaborative proposals from for-profit organizations and FFRDCs as well, but those organization can only be subawardees and not the lead awardee. An individual can only serve as PI or Co-PI in at most one submission, but may participate as senior personnel in multiple proposals.

Since an institution can only submit ONE lead proposal, please make sure you request a Letter of Collaboration from a Hub in advance, and work closely with your grant office.

Slide 21:

To become a Spoke, potential PIs should coordinate with a Big Data Hub to make sure that they align with the broader interests in the region and by its Big Data stakeholders. The regional BD Hub Steering Committee will provide general guidance to BD Spokes and assist with coordinating with the national BD Hub network, other BD Spokes, and the broader innovation ecosystem. We leave it to the spoke and the hub to articulate the proposal and the corresponding letter of collaboration, the details of how the Spoke and Hub will interact on a logistical, day-to-day level, and the details of how that collaboration will move forward.

As such, all proposals to this solicitation must include a letter of collaboration from a BD Hub coordinating institution. **Any proposals not including a letter of collaboration from a BD Hub coordinating institution**

In addition, each Hub has its guidelines on requesting a Letter of Collaboration.

Slide 22:

Each Hub has a distinct process for communicating with parties interested in submitting a BD Spoke proposal. Guidelines to each Hub's specifics can be found from the link on this slide.

The deadline for requesting a Letter of Collaboration from a Hub is June 19; this deadline is set to enable the Hubs to move forward efficiently. Please take a look as soon as possible; given the level of interest and timeline.

Slide 23:

A key objective of the BD Hubs and Spokes program is to build partnerships. Therefore, if your proposal expands or augments previously funded BD Spokes, the project team should coordinate in advance with those awarded BD Spoke teams. In this case, a Letter of Collaboration from a current BD Spoke is encouraged.

If you propose a SMALL or MEDIUM project that is substantively similar to that of a previously funded BD Spoke, even if the activity will take place in a different BD Hub region, the proposal must justify the need to fund the new activity. For example, the proposal can explain how the new activity might expand or augment the previously funded activities or differentiate itself from those other activities.

Slide 26:

There are a couple of differences between the 2016 and the 2017 BD Spoke solicitations:

- The 2017 solicitation calls for SMALL and MEDIUM proposals rather than Spokes and Planning Grants
- We anticipate to fund 10 to 20 awards, totaling $10M
- The full proposal deadline is Sept 18, 2017
- Letter of Intent is NOT required for the 2017 BD Spoke submission
- The 2017 solicitation is a limited competition as I mentioned earlier
- Finally, tangible outcomes are expected for the MEDIUM awards

Slide 27:

So both NSF and the Big Data Hubs are excited to see what your ideas are for Big Data Spokes.

Here is my contact information (fzhao@nsf.gov, 703‑‑292‑‑7344), and feel free to reach out to me if you have any questions. We'll open up the line to take some questions.