Good afternoon. We're Vipin Chaudhary, Amy Walton and Rajiv Ramnath from the NSF Office of Advanced Cyberinfrastructure, or OAC.

We are the Program Directors managing the **Cyberinfrastructure for Sustained Scientific Innovation** program, or CSSI.

In this webcast, we will provide a brief overview of the CSSI program, and describe some of the most important things you need to know about submitting a proposal.

Joining us to welcome you is Amy Friedlander, Deputy Director of the Office of Advanced Cyberinfrastructure at NSF.

Thank you.

Good afternoon. On behalf of the Office of Advanced Cyberinfrastructure, the Directorate for Computer and Information Science and Engineering, and the National Science Foundation, let me welcome you to this webinar. The program that will be described is an important step for us as we begin to develop a vision of integrated cyberinfrastructure. The solicitation represents the hard work of program directors across all of the participating directorates. So let me pause and thank them all.

The solicitation is only the next step. Now we want to hear from you. Today, you will ask your questions. But soon, we hope you will respond with your proposals and ideas. In those responses, we – and you – will learn more about what we mean by "integrated cyberinfrastructure", at least as it related to data and software; what we should do next; and how we can move forward toward building an advanced cyberinfrastructure in support of science.

Thank you all, in advance, for your time and attention. Now, let me turn this back over to the organizers.

Thank you Amy.

This webinar in intended to orient the research community to the CSSI competition, summarize the program and peer-review criteria, and answer questions. Of course, the ultimate goal is to improve the quality of your proposals.

Here is an outline of today's presentation. We'll start with a description of the CSSI program followed by an overview of the NSF 18-531 solicitation.

We will then take questions from you, the audience. Some of the questions we have already received are included at the end of the presentation.

This document will be available on the program website.

Next we talk about the CSSI program, its priorities and goals and how we implement it.

The CSSI program focuses on supporting robust, reliable and sustainable data and software cyberinfrastructure that will support and advance sustained scientific innovation and discovery. Thus, proposals are strongly encouraged to describe their approach to data management and quality software development through a defined software engineering process that includes software testing, the appropriate use of analysis tools and capabilities.

The CSSI program integrates the data and software elements of advanced cyberinfrastructure.

By integrating two major and long-running NSF program solicitations [Data Infrastructure Building Blocks (DIBBs) and Software Infrastructure for Sustained Innovation (SI2)] under a single umbrella called Cyberinfrastructure for Sustained Scientific Innovation (CSSI), NSF seeks to enable funding opportunities that are flexible and responsive to the evolving and emerging needs in integrated data and software cyberinfrastructure. This integration also minimizes multiple and overlapping submissions while encouraging integrated science-driven cyberinfrastructure.

The CSSI program is guided by six principles. The project must explicitly address these principles, which translate into solicitation-specific criteria.

- The project must be Science-driven, promoting science excellence, enabling fundamentally new scientific advances; and benefiting science and engineering communities beyond the participating communities
- *The project must be Innovative*, emphasizing unique NSF contributions; building the capability, capacity, and cohesiveness of a national CI ecosystem; and considers both the human and technical aspects of the CI.
- The project must be Collaborative, fostering partnerships and community development; actively engages CI experts, specialists and scientists working in concert with domain scientists who are users of CI.
- The project must be Leveraged, building on existing, recognized capabilities.
- The project must be Strategic, with management plans and metrics that encourage measurement of progress and sharing of results.
- *The project must be Sustained*, providing benefits beyond the participants and the lifetime of the award, and resulting in widely accessible long-term community cyberinfrastructure.

**Cyberinfrastructure for Sustained Scientific Innovation** is a crosscutting program that involves program officers from every NSF Directorate. Participating divisions and program officers are listed here, and several of our colleagues are attending today's webinar.

The participant list can also be reviewed on solicitation web page at: <a href="https://www.nsf.gov/pubs/2018/nsf18531/nsf18531.htm">https://www.nsf.gov/pubs/2018/nsf18531/nsf18531.htm</a>

The CSSI solicitation is responsive to national and NSF-wide objectives, as well as priorities in specific science areas. This slide identifies a national initiative and an NSF initiative that are of interest to this solicitation. The next several slides summarize areas of interest within the Foundation.

This CSSI solicitation welcomes proposals that advance the objectives of the National Strategic Computing Initiative (NSCI), an effort aimed at sustaining and enhancing the U.S. scientific, technological, and economic leadership position in high-performance computing (HPC) research, development, and deployment. Information about the NSCI together with the strategic plans, results of community workshops, background studies and other relevant resources, which suggest priority areas in both the domain sciences and the HPC and software infrastructure, are available at <a href="https://www.nsf.gov/nsci/">https://www.nsf.gov/nsci/</a>.

This CSSI solicitation also welcomes proposals that advance the objective of Harnessing the Data Revolution, one of NSF's Big Ideas. This initiative is aimed at fundamental data science research, research data cyberinfrastructure, and the development of a 21st century data-capable workforce. HDR will enable new modes of data-driven discovery – allowing researchers to ask and answer new questions in frontier science and engineering, generate new knowledge and understanding, and accelerate discovery and innovation.

Proposers are encouraged to review these materials for priority areas identified by the research community.

Within NSF, various organizations have additional specific information about their participation in this program:

The Office of Advanced Cyberinfrastructure (OAC) manages the CSSI program, and is especially interested in proposals that:

- Enable new science and engineering not previously possible.
- Contain innovation as an integral component of the project. Such research might encompass reproducibility, provenance, effectiveness, usability, and product adoption, adaptability to new technologies and to changing requirements, and the data and software development lifecycle processes used in the project;
- Build on existing community CI services and software, and leverage or complement other community cyberinfrastructure (CI) projects
- And seek to develop, deploy and sustain foundational infrastructure components, and interdisciplinary and omni-disciplinary computational tools and components.

The Directorate for Biological Sciences (BIO) is primarily interested in the CSSI program as a means to collaborate with other NSF directorates to support proposals that impact a multidisciplinary community that includes BIO-supported researchers. PIs wishing to submit projects that focus primarily on biological sciences should submit to the Advances in Biological Informatics program (ABI; see NSF 15-582).

The foundational research divisions within the **Directorate for Computer and Information Science and Engineering** (CISE) -- Computing and Communication Foundations (CCF), Computer and Network Systems (CNS), and Information and Intelligent Systems (IIS) -- are interested in software or data engineering and infrastructure projects that support research in all areas that sustain progress in CISE research areas or that advance and adapt CISE research to impact the data and software sustainability needs of other scientific disciplines. Please see the division-level descriptions in the solicitation for complete details.

The **Directorate for Education and Human Resources** (EHR) is interested in fostering novel, transformative, multidisciplinary approaches that address the use of large data sets and/or learning analytics to create actionable knowledge for improving STEM teaching and learning environments (formal and informal) in the medium term, and to revolutionize learning in the longer term.

The Directorate of Engineering seeks proposals for innovative software and data infrastructure that enable major advances in fundamental research funded by the divisions. ENG will support proposals that give the engineering research community broad and sustained access to HPC and data and software platforms and technologies that support emerging research opportunities. The goal is to broaden the use of advanced computing and data by the ENG research community to enhance research productivity and open new pathways to discovery. Please see the division-level descriptions in the solicitation for complete details.

The Directorate for Geosciences (GEO) is interested in the following research fields: atmospheric and geospace science, earth science, ocean science, and polar science. The directorate welcomes proposals that focus on the development and implementation of automated, executable, and optimized workflows that:

- are capable of real- and near-real-time archiving and manipulation of sensor and other field-based data, including experimental and/or simulation data;
- promote seamless discovery, access, and transfer of data and metadata across data resources and centers that are supported by GEO;
- "leverage-through-sharing" of existing investments in university, federal, and commercial computing and infrastructure;
- engage community models for the assimilation and use of data for initialization, state estimation, or sensitivity analysis; and
- encourage the development or reuse of computational techniques (i.e., high-performance distributed computing, machine learning, cloud computing, etc.) without which simulations would be difficult to conduct or large volumes of data could not be manipulated, analyzed, and transferred. The proposed computational techniques must be framed as sub-tasks within the proposed workflow.

Please see the division-level descriptions in the solicitation for complete details.

The Directorate for Mathematical and Physical Sciences (MPS) appreciates that software and data cyberinfrastructure enable scientific advances and discovery across MPS. In all MPS disciplines, there is need for cyberinfrastructure to support innovative scientific inquiry based on software and data that are findable, accessible, reusable, provenance traceable, and sustainably maintainable. Data cyberinfrastructure may additionally combine the elements of algorithms, software, computation, networks, task automation, or custom hardware to support data-centric approaches to MPS science. Data may be derived from experiment, observation, or computation, and may be diverse in kind consistent with science across MPS. Please see the division-level descriptions in the solicitation for complete details.

The **Directorate for Social, Behavioral, and Economic Sciences** (SBE) is interested in proposals that support the Directorate's research priorities, such as those outlined in SBE 2020 (<a href="https://www.nsf.gov/sbe/sbe\_2020/">https://www.nsf.gov/sbe/sbe\_2020/</a>). SBE is particularly interested in using CSSI to support projects building on other infrastructure activities such as Metadata for Long-standing Large-Scale Social Science Surveys (META-SSS) (<a href="https://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504705">https://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=504705</a>) and Resource Implementations for Data Intensive Research in the Social, Behavioral and Economic Sciences (RIDIR) (<a href="https://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=505168&org=SES&from=home">https://www.nsf.gov/funding/pgm\_summ.jsp?pims\_id=505168&org=SES&from=home</a>). SBE also welcomes innovative approaches to big data problems in SBE-focused domains consistent with NSF's Big Idea of Harnessing the Data Revolution. SBE encourages proposals that further the goals of SBE and at least one other participating NSF directorate.

Next we talk about this specific solicitation opportunity (NSF 18-531), including classes of investments, PI eligibility, and review criteria.

The CSSI umbrella includes four different classes of investment. Only the first two investment classes are included in solicitation NSF 18-531; the other classes will not be covered in this webinar.

First is the **Elements.** This class of investment targets small groups that will create and deploy robust capabilities for which there is a demonstrated need that will advance one or more significant areas of science and engineering.

The next class of investment is the **Framework Implementations** that targets larger, interdisciplinary teams organized around the development and application of common infrastructure aimed at solving common research problems faced by NSF researchers in one or more areas of science and engineering, resulting in a sustainable community framework serving a diverse community or communities.

The third class of investment is the **Planning Grants for Community Cyberinfrastructure** where the focus is on the establishment of long-term capabilities in cyberinfrastructure, which would serve a research community of substantial size and disciplinary breadth.

Finally, the fourth class of investment is the **Community Cyberinfrastructure Implementations** that focuses on the establishment of long-term hubs of excellence in cyberinfrastructure and technologies, which will serve a research community of substantial size and disciplinary breadth.

For the current solicitation NSF 18-531, the proposers are asked to identify whether their proposal is an "Elements" or "Framework" proposal in the proposal title. Proposers are also asked to identify whether the proposal is a "Data" proposal or a "Software" proposal within the title, based on whether the proposed cyberinfrastructure will primarily support data-driven research or software-driven research. Thus, there are four classes of proposals and investments for this solicitation. However, OAC recognizes that proposal submissions will span a continuum of cyberinfrastructure possibilities. Therefore, researchers with questions as to how their proposal should be classified should not hesitate to contact the cognizant program officers listed in this solicitation.

For proposers who are familiar with the DIBBS and SI2 programs, this slide shows the relationship between CSSI and DIBBS and SI2. This information is only for familiarity and adds no other value.

The DIBBS pilot demonstrations are now CSSI data elements. The SI2 SSE are now CSSI software elements. The DIBBS early implementations are now CSSI data frameworks. The SI2 SSI are now CSSI software frameworks.

The SI2 software conceptualizations are now CSSI planning grants for community software cyberinfrastructure. The SI2 software innovation institutes are now CSSI community software cyberinfrastructure implementations.

The Planning Grants for Community Data Cyberinfrastructure and the Community Data Cyberinfrastructure Implementations were not part of DIBBS.

Please note that this solicitation, NSF 18-531 is only targeting the Elements and Framework Implementations Investment class and we will only address details and questions regarding those in this webinar.

The data and software elements awards shall not exceed a total of \$600,000 and 3 years duration.

Framework Implementations awards shall range from \$200,000 to \$1M per year, and shall be 3 to 5 years in duration.

Projects in the upper portion of this range must be exceptional in terms of scientific impact, and as with all proposals, should be discussed with program officers from the divisions that fund the researchers that would be impacted. Proposed funding amounts should be commensurate with the work being proposed, the size of the community that will be affected, and the level of impact anticipated.

Up to 15 Element awards, and 13 Framework Implementations awards are anticipated, subject to the availability of funds

With an anticipated total funding for \$34M, up to \$9,000,000 is expected to be available for Elements awards, and up to \$25,000,000 is expected to be available for Framework Implementations awards, subject to the availability of funds.

All proposals to this solicitation, namely, the data and software elements and the data and software framework implementations have the same deadline - April 18, 2018.

We expect the review process lasting from May to July of this year and anticipate making announcements of awards in Fall 2018.

The eligibility criteria for the CSSI program are as follows:

Proposals may only be submitted by universities and colleges or non-profit, non-academic organizations. Federally-funded research and development centers (FFRDCs) may not receive funds directly from NSF under this solicitation.

The number of proposals per principal investigator, co-principal investigator, or senior personnel is limited to one. An individual may participate in a proposal as principal investigator, co-principal investigator, or other senior personnel in at most one proposal in a given calendar year across all categories of proposals. In the event that any individual exceeds this limit, any proposal submitted to this solicitation with this individual listed as PI, co-PI, or Senior Personnel after the first proposal is received at NSF will be returned without review. No exceptions will be made

Please review the solicitation for details.

Some important aspects to note about the CSSI Cover Sheet.

For the NSF Unit of Consideration, The "Divisions" section should automatically be selected. Select one class from the drop-down list in FastLane as the program(s) to consider the proposal. For example, Data Elements or Data Frameworks proposals should choose "DATANET" as the Program. Software Elements or Software Frameworks should choose "Software Institutes" as the Program. Grants.gov users should refer to Section VI.1.2. of the NSF Grants.gov Application Guide for specific instructions on how to designate the NSF Unit of Consideration.

For the proposal title, provide a short informative title for the proposed project. To assist NSF staff in sorting proposals for review, proposal titles should begin with "Elements:" or "Framework:", followed by "Data" or "Software". Proposals that are to be considered as responsive to the NSCI should additionally prefix their title with "NSCI". Proposals that are to be considered as responsive to HDR should additionally prefix their title with "HDR".

An example title would be Element:Data:HDR:MyProjectTitle

A second example title would be Element:Software:MyProjectTitle indicating that HDR or NSCI are not required prefixes and will be used only if the proposals address priorities related to HDR or NSCI.

A third example title would be Frameworks:Software:NSCI:MyProjectTitle

Additional documents include:

A data management plan and postdoctoral trainee mentoring plan (if the project includes such trainees). This is a standard NSF requirement. CSSI reviewers will pay close attention to the data management plan.

For framework implementation proposals, a management and coordination plan is also required. The specific roles of the principal investigators, co-principal investigators, other senior personnel, and paid consultants at all institutions involved must be outlined. Also, there must be a description of how the project will be managed across institutions and disciplines, identification of the specific coordination mechanisms that will enable cross-institution and/or cross-discipline scientific integration, and pointers to the budget line items that support these management and coordination mechanisms.

**Letters of Collaboration (if any) should** include documentation of funded or unfunded collaborative arrangements of significance to the proposal. Letters of collaboration should be limited to stating the intent to collaborate and should not contain endorsements or evaluation of the proposed project. The REQUIRED format for letters of collaboration is in the NSF Proposal & Award Policies & Procedures Guide (PAPPG).

Additional documents also include:

**Project Personnel and Partner Institutions.** These are required for all award categories. You must provide current, accurate information for all personnel and institutions involved in the project. NSF staff will use this information in the merit review process to manage conflicts of interest. The list must include all PIs, Co-PIs, Senior Personnel, paid/unpaid Consultants or Collaborators, Subawardees, Postdocs, project-level advisory committee members, and writers of letters of support. See details in the solicitation.

Additional single copy documents include the Collaborators and Other Affiliations Information:

- Collaborators & Other Affiliations information specified in the PAPPG should be submitted using the spreadsheet template found at https://www.nsf.gov/cise/collab/. For each proposal, a completed spreadsheet for each PI, co-PI, or senior personnel must be uploaded directly into Fastlane in .xls or .xlsx format as a "Collaborator and Other Affiliations" Single Copy Document.
- NSF staff use this information in the merit review process to help manage reviewer selection; the spreadsheet will ensure the Collaborator and Other Affiliations information has a common, searchable format.

Note the distinction of this information in previous slide (additional documents(2)): the listing of all project participants is collected by the project lead and entered as a Supplementary Document, which is then automatically included with all proposals in a project. The Collaborators and Other Affiliations (this document) are entered for each participant within each proposal and, as Single Copy Documents, are available only to NSF staff.

See details in the solicitation

As for all proposals received by NSF, CSSI reviewers and panelists will be asked to consider the intellectual merit and broader impact for each proposal for their reviews, panel discussions, and panel summaries. In addition to these standard criteria, SI2 reviewers and panelists will also be asked to consider additional review criteria that are unique to the CSSI program. More on this in a few moments.

When evaluating NSF proposals, reviewers are asked to consider:

- what the proposers want to do?
- why they want to do it?
- how they plan to do it?
- how they will know if they succeed?
- what benefits would accrue if the project is successful?

These issues apply both to the technical aspects of the proposal (the intellectual merits) and the way in which the project may make broader contributions (the broader impacts).

In addition to the standard NSF review criteria, the proposals will be evaluated on CSSI-specific review criteria, namely,

- 1) To what extent is the proposed project science-driven? How will the project outcomes fill well-recognized science and engineering needs of the research community, and advance research capability within a significant area or areas of science and engineering? What will be the broader impacts of the project, such as, its benefits to science and engineering communities beyond its initial targets, under-represented communities, education and workforce development? The project description should provide a compelling discussion of the potential to benefit its intended as well as broader communities.
- 2) To what extent is the proposed project innovative? What innovative and transformational capabilities will the project bring to its target communities? How will the project integrate innovation and discovery into the project activities, such as through empirical research embedded as an integral component of the project activities. Such research might encompass reproducibility, provenance, effectiveness, usability, and adoption of the components, its adaptability to new technologies and to changing requirements, and the development lifecycle processes used in the project;
- 3) To what extent does the proposed project involve close collaborations among stakeholders? How will the project activities engage cyberinfrastructure (CI) experts, specialists and scientists working in concert with the relevant domain scientists who are users of CI.
- 4) To what extent does the proposed project build on existing, recognized capabilities? How will the project activities build on and leverage existing NSF and national cyberinfrastructure investments, as appropriate?
- 5) How well described are the project plans, and system and process architecture? The project description should include high-quality management plans. The project plan should include user interactions and a community-driven approach, and provide a timeline including a proof-of-concept demonstration of the key components. The proposal must include a list of tangible metrics to be used to measure the success of the project activities, and measure progress along the way. If the outcome of the project is software or data cyberinfrastructure, the architecture of the CI and the engineering process to be used for the design, development, documentation, testing, validation and release of the software, its deployment and associated outreach to the end user community, and an acceptance and evaluation plan that involves end users, all must be sufficiently described. The description of the CI architecture and processes should explain how security, trustworthiness, provenance, reproducibility, and usability will be addressed by the project and integrated into the proposed system and the engineering process, and how adaptability to new technologies and changing requirements will be addressed by the project and built into the proposed system, as appropriate.
- 6) How well does the project address the achievement of sustained and sustainable impacts? The project description should address how the project outcomes and its activities will have long-term impacts, and how these will be sustained beyond the lifetime of the award, as appropriate. If the outcome of the project is software or data cyberinfrastructure, the proposal should identify what license will be used for the released CI, and why this license has been chosen. PIs who have been previously funded under previous CI awards should show quantifiable evidence of the use, impact and sustainability of the previously funded work (and include a citation to the published CI in their biographical sketches as one of their relevant products, if appropriate).

A competitive CSSI proposal will Identify science and engineering challenges where the proposed cyberinfrastructure enables fundamental new science advances, and describe how the proposed project fosters partnerships and community development that will have a significant impact on science and engineering research.

The proposal will indicate how the proposed cyberinfrastructure builds capability, capacity and cohesiveness of a national CI ecosystem; and

Provide a compelling discussion of the cyberinfrastructure's potential use by a wider audience and its contribution to a national cyberinfrastructure.

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.

A key question – and a major change from prior solicitations -- concerns the limits on the number of proposals an individual may participate in under this solicitation (NSF 18-531).

- An individual may participate as PI, co-PI, or other Senior Personnel on at most one proposal across the Elements and Framework Implementations for this solicitation. Thus, if an individual participates on an Elements proposal, he or she may not participate on a Framework Implementations proposal, and vice versa.
- Note that any individual whose biographical sketch is provided as part of the proposal will be considered as Senior Personnel in the proposed activity, with or without financial support from the project.
- In the event that any individual exceeds this limit, any proposal submitted to this solicitation with this individual listed as PI, co-PI, or Senior Personnel after the first proposal is received at NSF will be returned without review.
- No exceptions will be made.

Proposals must be received by 5 p.m. submitter's local time on the established deadline date. Failure to submit by 5 p.m. submitter's local time will result in the proposal not being accepted.

OAC recognizes that software and data infrastructure constitute fundamental infrastructure that cross-cuts academic, government, civic, and commercial organizations. The program encourages proposals to explore novel partnerships beyond academe wherever beneficial and permissible within the guidelines of the NSF *Proposal & Award Policies & Procedures Guide (PAPPG)*.

- What types of organizations organizations are allowed to submit proposals?
  - *Universities and Colleges* Universities and two- and four-year colleges (including community colleges) accredited in, and having a campus located in, the US acting on behalf of their faculty members. Such organizations also are referred to as academic institutions.
  - *Non-profit, non-academic organizations*: Independent museums, observatories, research labs, professional societies and similar organizations in the U.S. associated with educational or research activities.
  - NSF-sponsored federally funded research and development centers (FFRDCs), provided that they are not including costs for which federal funds have already been awarded or are expected to be awarded.

#### The next question on

- How can other organizations participate?
  - Organizations eligible to serve as subawardees are all organizations eligible under the guidelines of the NSF *Proposal & Award Policies & Procedures Guide (PAPPG)*.

Here are a couple more questions about collaborations.

- How can a proposal integrate industry collaboration into the project?
  - Industry participants may be included as a subaward within the proposal.
  - Industry investigators may serve as co-PIs or senior personnel on a proposal. (See PAPPG, Part I, E.3).
  - Industry participants may be (unfunded) collaborators.
  - Industry participation should be integrated through the management plan.

The next question on

- Can a foreign organization submit a proposal?
  - NSF rarely provides support to foreign organizations. NSF will consider proposals for cooperative projects involving US and foreign organizations, provided support is requested only for the US portion of the collaborative effort.

The next three questions address the differences between this and other solicitations.

- First, what is the difference between a data and a software proposal submission?
  - The data proposal (either a Data Element or a Data Framework proposal) has data attributes as the core enabler for the success of the project.
  - The software proposal (either a Software Element or a Software Framework proposal) has software attributes as the core enabler for the success of the project.
  - There are obvious areas of overlap; proposers are encouraged to clearly identify where they expect to make major contributions.
- Second, how do CSSI proposals differ from Computational and Data-Enabled Science and Engineering (CDS&E) proposals?
  - CDS&E emphasizes research in, rather than the development of, cyberinfrastructure systems. CDS&E also has a limited set of participating NSF Directorates.
  - CSSI includes a broader range of Directorates, and focuses upon development of data and software systems that support research.
  - And finally, how are data proposals to CSSI different from BIGDATA proposals?
  - Data proposals to CSSI focus upon innovative, use-inspired and user tested infrastructure that contributes to future discovery across communities.
  - BIGDATA focuses on research challenges in the foundations of data science, and development of innovative applications.

On behalf of the National Science Foundation and the CSSI team, we thank you for participating in this webinar.

The slides and the script for this webcast, as well as an audio recording, will be available at http://www.nsf.gov/events/. On that page, you'll need to look for this webcast among the list of events. I invite your questions now, via email or via telephone to Vipin Chaudhary, Amy Walton or Rajiv Ramnath. You can also find contact details for program officers from other NSF Directorates who are involved in the CSSI program on the solicitation web page.