Harnessing the Data Revolution (HDR) Information Webinar

February 15, 2019,

1:00 pm – 3 pm Eastern

Questions: Email

HDR-DIRSE@nsf.gov

Agenda

- Introduction
- Welcome from NSF Leadership

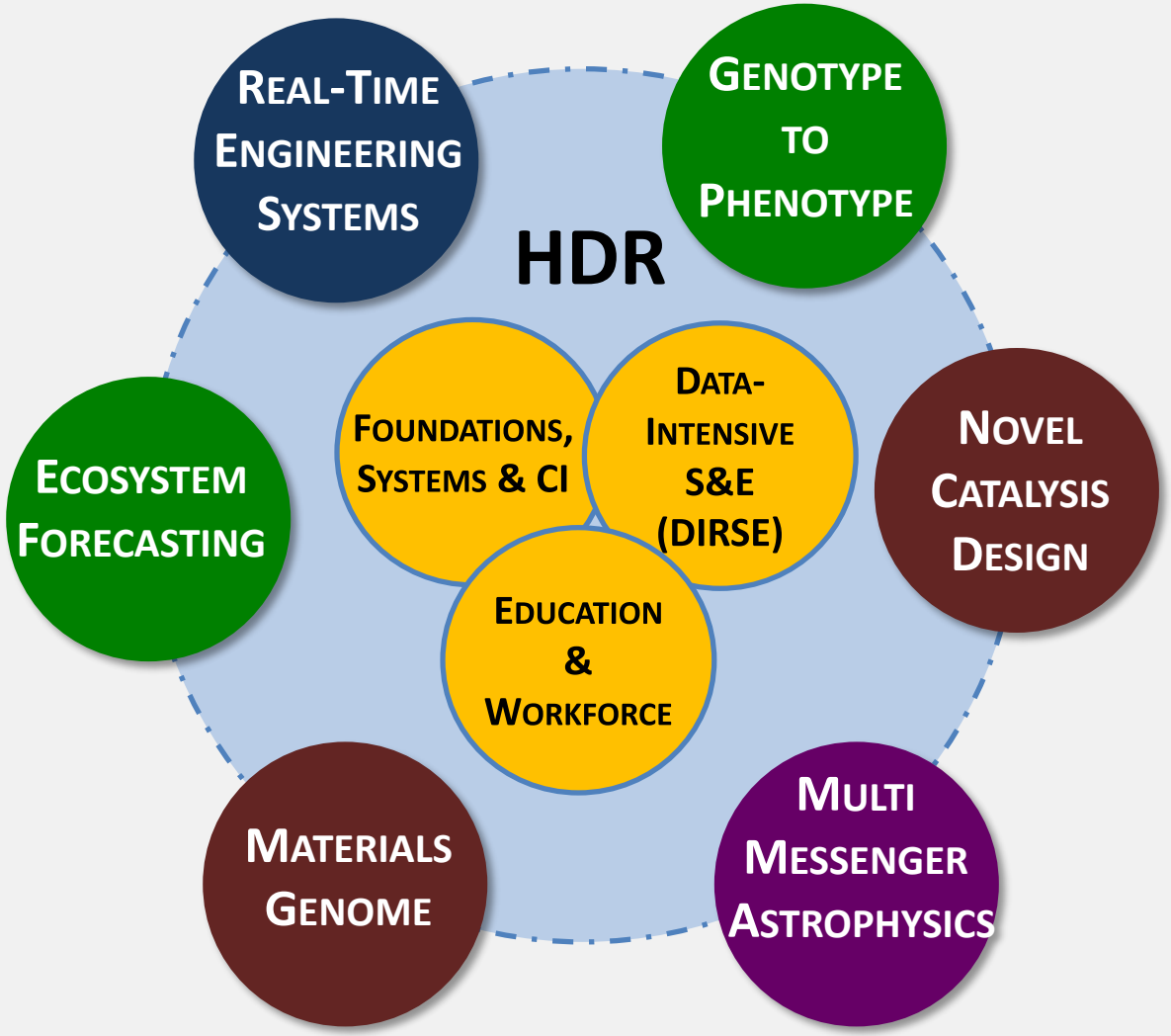
  Jim Kurose (CISE) and **Anne Kinney (MPS)**
- Overview of the Harnessing the Data Revolution (HDR) Big Idea

  HDR Co-Chairs: **Juan Meza (MPS/DMS)** and **Manish Parashar (CISE/OAC)**
- FY19 HDR Activities
- Solicitations: **Goals, Proposal Preparation and Review Criteria (Prerecorded)**
  - Ideas Labs: **Nandini Kannan (MPS/DMS)**
  - Frameworks: **Amy Walton (CISE/OAC)**
  - TRIPODS: **Tracy Kimbrel (CISE/CCF)**
- Q&A

*"Engage NSF's research community in the pursuit of **fundamental research in data science and engineering**, the development of a cohesive, federated, national-scale approach to **research data infrastructure**, and the development of **a 21st-century data-capable workforce**."*

HARNESSING THE DATA REVOLUTION VISION

HDR will enable new modes of data-driven discovery that will allow fundamental questions to be asked and answered at the frontiers of science and engineering



Overview of HDR Funding Opportunities

- **Data-Intensive Research in Science and Engineering: Ideas Lab (Nandini Kannan)**

  [NSF 19-543: Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Ideas Labs (I-DIRSE-IL)]

- **Data-Intensive Research in Science and Engineering: Frameworks (Amy Walton)**

  [NSF 19-549: Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Frameworks (I-DIRSE-FW)

- **Foundations: Transdisciplinary Research in Principles of Data Science Phase I (Tracy Kimbrel)**

  [NSF 19-550 : Harnessing the Data Revolution (HDR): Transdisciplinary Research in Principles of Data Science Phase I]

- **EDUCATION AND WORKFORCE: (Deadline: February 14, 2019)**

  [NSF 19-518: Harnessing the Data Revolution (HDR): Data Science Corps (DSC)]

Data-Intensive Research in Science and Engineering (DIRSE) Institutes: Goal

- Create an **integrated** fabric of interrelated institutes that can accelerate discovery and innovation in multiple areas of data-intensive science and engineering.
- Support **convergence** between science and engineering research communities as well as expertise in data science foundations, systems, applications, and cyberinfrastructure.
- Enable breakthroughs in science and engineering through **collaborative, co-designed** programs to formulate innovative data-intensive approaches to address critical national challenges.

HDR Institutes Roadmap

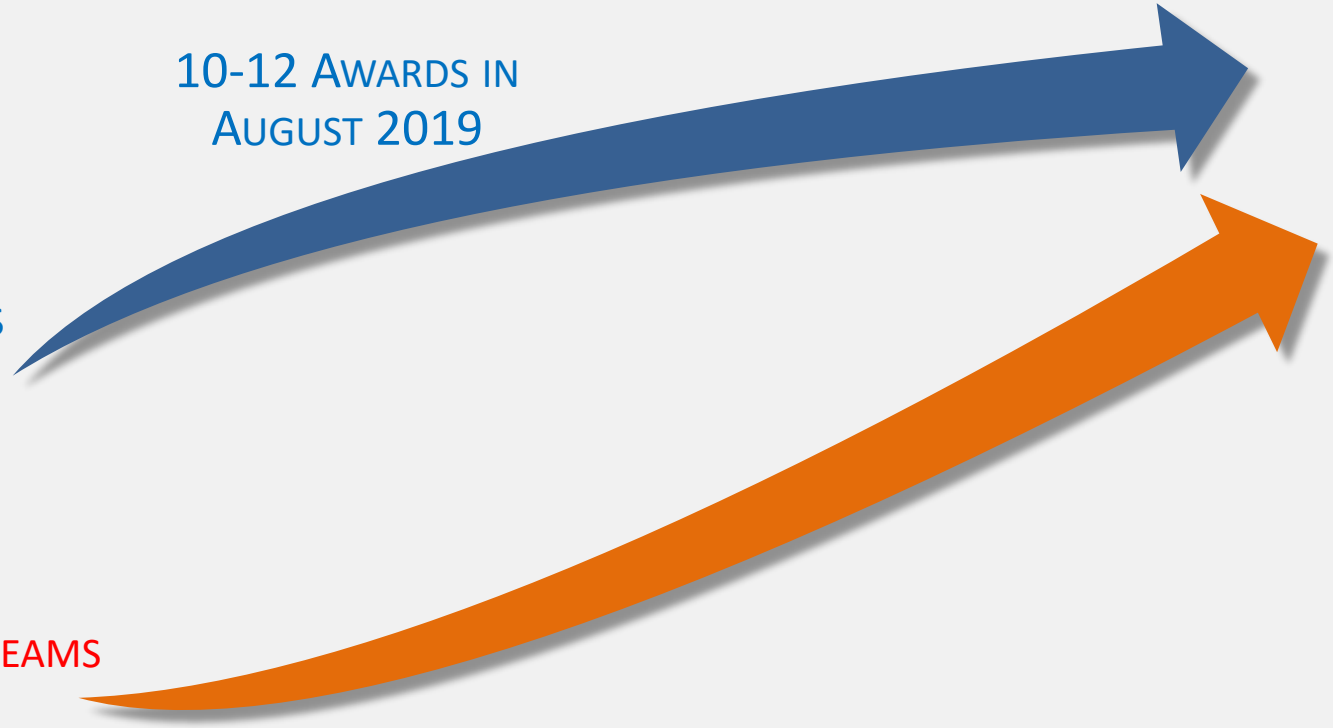
Conceptualization

10-12 Awards in August 2019

New Teams formed via Ideas Labs

Existing Teams pursue Frameworks

8-10 awards in August 2019



2021 Timeframe

Open Competition for 4-5 institutes that bring together multiple science and engineering communities

FY 19 HDR SOLICITATIONS

- **NSF 19-543**: Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Ideas Labs (I-DIRSE-IL)

  Nandini Kannan (MPS/DMS)

- **NSF 19-549**: Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Frameworks (I-DIRSE-FW)

  Amy Walton (CISE/OAC)

- **NSF 19-550**: Harnessing the Data Revolution (HDR): Transdisciplinary Research in Principles of Data Science Phase I

  Tracy Kimbrel (CISE/CCF)

DIRSE - IDEAS LAB (NSF 19-543)

PRELIMINARY PROPOSAL DEADLINE: MARCH 4, 2019

IDEAS LABS: May 20-24, 2019

Data-Intensive Research in Science and Engineering

Focus on **science and engineering areas** that:

1. are at a "**tipping point**" where a timely investment in data-intensive approaches has the maximum potential for a transformative effect,
2. have needs that can benefit from **interdisciplinary investments in data analytics infrastructure**, and
3. represent **investment priorities** for the participating NSF directorates during, and beyond, the lifetime of the HDR Big Idea

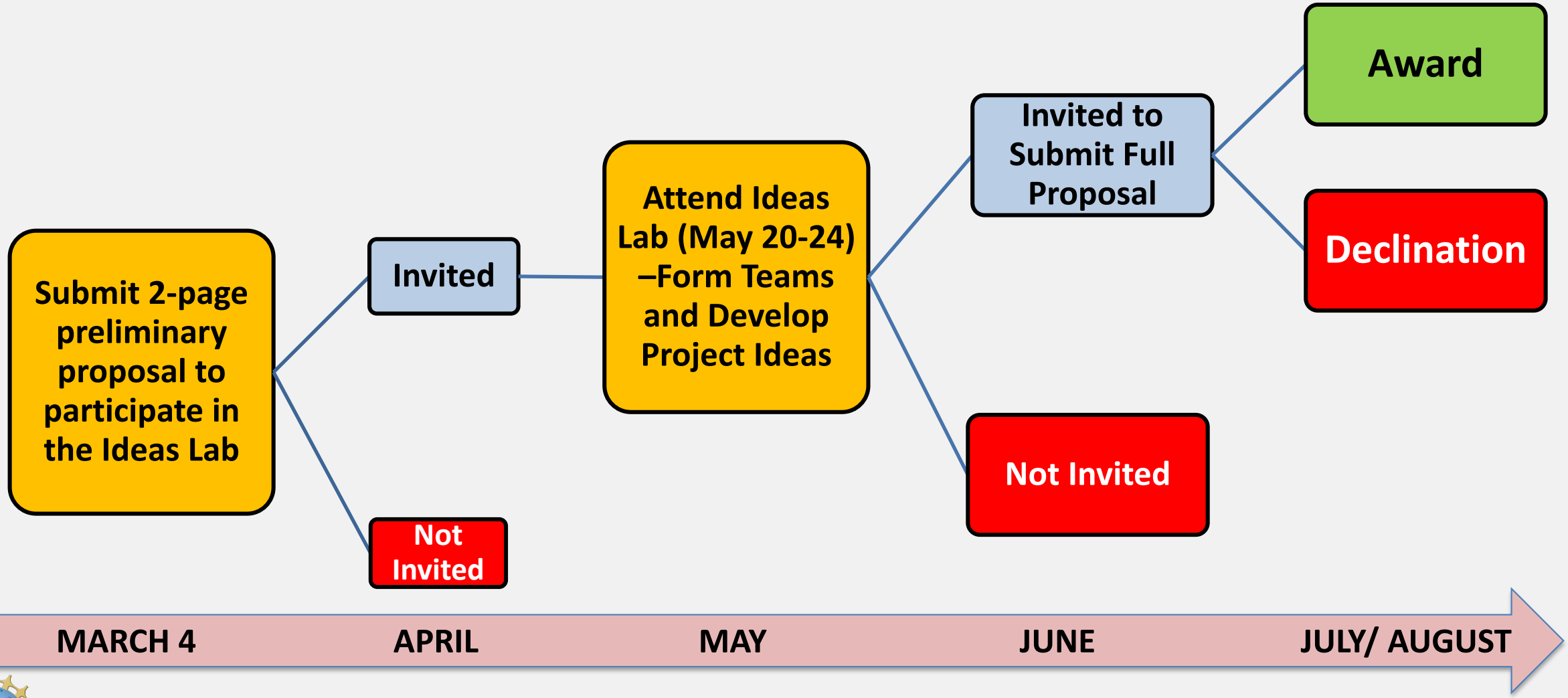
DIRSE - Ideas Labs

Goal: Foster convergent approaches to enable data-intensive research in science and engineering through a series of facilitated activities.

Mechanism: Recruit participants from both the research (all science and engineering disciplines) and technical (computer and computational science, mathematics, statistics, and information science) domains.

Create an environment for interdisciplinary teams, comprising **individual participants with complementary expertise, to emerge organically through an iterative process.**

IDEAS LAB: PROCESS



Ideas Labs Process

- **Preliminary proposals** (2 page conforming to specified guidelines) for participating in the Ideas Labs only via Fastlane.
- Participation in an Ideas Lab is **required to be eligible to submit a full conceptualization proposal** pursuant to this solicitation.
- Multidisciplinary ideas developed in an Ideas Lab will be submitted as full conceptualization proposals to NSF **by invitation only.**

Preliminary Proposals for Ideas Labs

- Submission of **Preliminary Proposals** is **required** for participation in the Ideas Labs and will be considered an **indication of availability to attend and participate** through the full course of a five-day Ideas Lab (location near NSF headquarters in Alexandria, VA, May 20-24, 2019
- Preliminary proposal must come from one individual and **cannot include co-PIs or collaborators**.
- Participants in the Idea Labs will be selected on the basis of information submitted in the preliminary proposal

Deadline for Preliminary Proposals: MARCH 4, 2019

Ideas Labs: May 20-24, 2019

- **Intensive five-day residential workshops** focused on finding innovative and bold transdisciplinary solutions to grand challenge problems.

- Development of multidisciplinary collaborative proposals through a real-time and iterative review process involving teams of experts (mentors).

Following the Ideas Labs, teams of researchers with the most potentially transformative ideas will be invited to submit **full proposals**

DIRSE - FRAMEWORKS

DUE DATE: MAY 7, 2019

Frameworks Goal

- Enable **interdisciplinary teams** to **conceptualize and pilot** new modes for engaging in data-intensive science and engineering
- Build **innovative connections** between scientific groups and data scientists and engineers, to integrate research infrastructure and education infrastructure.

FRAMEWORKS: Attributes

1. An **innovative vision** for an Institute that articulates the potential for **transformative outcomes in science and engineering** through data-intensive research;
2. **Collaboration**s among domain scientists, data scientists, and cyberinfrastructure experts that **leverage of existing research infrastructure and resources**;
3. A **vision for transforming communities of practice and the future workforce in data-intensive research**. These activities must be well-integrated into the Framework research activities;

FRAMEWORKS: Attributes

4. The **development of convergent modes of collaboration that cross institutional boundaries** and that can expand to engage new collaborators throughout the lifecycle of a future Institute. Approaches must be scalable and extensible for a future Institute; and
5. A **common means of openly sharing outcomes** from the Institute, including but not limited to scientific and education outcomes, data, algorithms and models

Overall, proposals must describe a rationale that justifies the need for the **collective effort** of a group of domain scientists and data scientists to **enable transformative advances** in data-intensive research.

FRAMEWORKS: Eligibility

- **Proposals may only be submitted by:**
  - Institutions of Higher Education
  - Non-profit, non-academic organizations
  - NSF-sponsored federally funded research and development centers (FFRDCs)

- **Limit on Number of Proposals per PI/Co-PI/Senior Personnel: 1**
  - An individual may participate as Principal Investigator, co-Principal Investigator or other Senior Personnel in at most one Framework proposal pursuant to this solicitation
  - 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

Number of Awards and Funding Amounts

- **Estimated Number of Awards**: 8 to 10 awards in FY 2019 *pending availability of funds and the type, scale, and variety of project ideas proposed.*

- **Anticipated Funding Amount**: Up to a total of $21 million is available for 8 - 10 two-year awards.

Note: Estimated program budget, number of awards and average award size/duration are subject to the availability of funds.

DIRSE-FW PROPOSAL REVIEW General and Solicitation Specific Review Criteria

NSF Review Criteria

- **Intellectual Merit**: Encompasses the potential to advance knowledge

- **Broader Impacts**: Encompasses the potential to benefit society and contribute to the achievement of specific, desired societal outcomes.

DIRSE-FW Solicitation-Specific Review Criteria

1. How well does the proposed Framework provide a **clear vision for an Institute with potentially transformative outcomes** for data intensive research in science and engineering questions that address national priorities?
2. How do the **modes of collaboration** and the proposed team facilitate the goals of the Framework?
3. How well are existing resources and infrastructure **leveraged**?
4. What is the potential for the Framework to develop systems or cyberinfrastructure solutions that are **scalable and extensible to broader science and engineering questions**?

DIRSE ORGANIZATIONS AND CONTACT INFORMATION

Participating NSF organizations

- **Directorate for Biological Science (BIO)**
  - **Karen Cone, Division of Molecular & Cellular Biosciences**
  - **Peter H. McCartney, Division of Biological Infrastructure (DBI)**

- **Directorate for Computer & Information Science & Engineering (CISE)**
  - **James Donlon, Division of Information & Intelligent Systems**
  - **Tracy Kimbrel, Division of Computing and Communication Foundations**
  - **Sylvia Spengler, Division of Information & Intelligent Systems**
  - **Amy Walton, Office of Advanced Cyberinfrastructure**

Participating NSF organizations

- **Directorate for Education & Human Resources (EHR)**
  - John C. Cherniavsky, Division of Research on Learning in Formal and Informal Settings

- **Directorate for Engineering (ENG)**
  - Anthony Kuh, Division of Electrical, Communications & Cyber Systems
  - Alexis Lewis, Division of Civil, Mechanical & Manufacturing Innovation
  - Triantafillos J. Mountziaris, Division of Chemical, Bioengineering, Environmental, and Transport Systems

Participating NSF organizations

- **Directorate for Geosciences (GEO)**
  - **Eva Zanzerkia, Division of Earth Sciences**

- **Directorate for Mathematical & Physical Sciences (MPS)**
  - **Lin He, Division of Chemistry**
  - **Daryl W. Hess, Division of Materials Research**
  - **Nandini Kannan, Division of Mathematical Sciences**
  - **Vyacheslav (Slava) Lukin, Division of Physics**
  - **Nigel Sharp, Division of Astronomical Sciences**

Participating NSF organizations

- **Directorate for Social, Behavioral & Economic Sciences (SBE)**
  - Cheryl L. Eavey, Division of Social and Economic Sciences
  - Larry Gottlob, Division of Behavioral and Cognitive Sciences

HDR Transdisciplinary Research in Principles of Data Science (TRIPODS) (NSF 19-550)

Letter of Intent Deadline: March 25

Full Proposal: April 24 - May 8, 2019

TRIPODS OVERVIEW

- Collaboration between the Division of **Computing and Communication Foundations** (CCF) in the Directorate for Computer & Information Science & Engineering (CISE), the **Division of Mathematical Sciences** (DMS) in the Directorate for Mathematical and Physical Sciences (MPS), and the Division of **Electrical, Communications and Cyber Systems** (ECCS) in the Directorate for Engineering
- Focuses on the theoretical foundations of data science, --**core algorithmic, mathematical, and statistical principles**.

TRIPODS PROGRAM GOAL

Bring together communities from **electrical engineering, mathematics, statistics, and theoretical computer science** to develop the **theoretical foundations of data science** through institutes for **integrated research and training activities**.

TRIPODS PHASE I

- Development of **small collaborative Institutes** that will bring together the four disciplines.
- Address **fundamental research and training** in the theoretical foundations of data science, and describe the significant involvement of at least three of the four communities.

TRIPODS PHASE I

- Teams will develop capacity and demonstrate the ability to scale activities for full Institute operations by operating as smaller Institutes
- Novel approaches encouraged
- Traditional center-like activities are anticipated
  - workshops
  - training of students & postdocs
  - workforce development
  - community building

TRIPODS PHASE I → II

- 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 TRIPODS program

- Barriers related to **different terminology and formalisms for overlapping concepts and methods** used by different communities
- **Relevance to application domains and industry**
  - Addressing other Big Ideas encouraged
  - Partnerships encouraged
- **Entire "data to knowledge to action"** pipeline, including dynamic data collection
- Foundational tools for **data sets not handled by present tools**
- **Unified curricula** for data science: addressing experimental validation, ethical behavior, and interdisciplinary communication skills

POSSIBLE TRIPODS 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

POSSIBLE TRIPODS RESEARCH FOCI

- 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

TRIPODS : Eligibility

- **Proposals may only be submitted by:**
  - Institutions of Higher Education

- **Limit on Number of Proposals per PI/Co-PI/Senior Personnel: 1**
  - 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.

TRIPODS : Eligibility Restrictions

- **INSTITUTION:** Any institution that received an award under the previous TRIPODS Phase I solicitation NSF 16-615 (see https://www.nsf.gov/news/news_summ.jsp?cntn_id=242888) 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.
- **PI/ Co 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

TRIPODS Number of Awards and Funding Amounts

- **Estimated Number of Awards**: Fourteen to twenty awards of up to $1,500,000 per award are anticipated. The number of awards will be subject to availability of funds and receipt of proposals of adequate quality.

- **Anticipated Funding Amount**: $22,000,000

- Proposers may request up to $500,000 per year for a duration of three years.

Note: Estimated program budget, number of awards and average award size/duration are subject to the availability of funds.

TRIPODS PROPOSAL REVIEW
Solicitation Specific Review Criteria

TRIPODS: SOLICITATION SPECIFIC REVIEW CRITERIA

a) Describe a **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

b) "Broad themes of the program" listed in the Program Description

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

d) Transdisciplinarity/Synergy:

- Bringing together theories and approaches from electrical engineering, mathematics, statistics, and theoretical computer science
- Synergy between the groups

e) Vision: Ability to identify and articulate a vision for the foundations of data science

TRIPODS: SOLICITATION SPECIFIC REVIEW CRITERIA

f) **Quality and Value of Collaboration**:
   - Project expertise is **complementary, and well-suited** to the program goals
   - **Specific roles** of each collaborating investigator are made clear
   - **Collective team's expertise** represents at least three of the four communities

g) Well-developed plan for **communication and interaction with the domain sciences and industry**

h) Clear plan and rationale for an investment of the size proposed, **including clear plans to develop capacity for potential future Phase II operations**

i) Plan for Collaboration and Evaluation: **measures of success**, both for Phase I operations and **development of capability for a potential Phase II**

j) Clear plan for **thoughtful, ongoing assessment of all Institute activities**

Participating NSF organizations

Directorate for Computer & Information Science & Engineering, Division of Computing and Communication Foundations (CISE/CCF)

- Tracy Kimbrel
- Phillip Regalia
- Rahul Shah

Directorate for Engineering, Division of Electrical, Communications & Cyber Systems, (ENG/ECCS)

- Anthony Kuh
- Akbar Sayeed

Directorate for Mathematical & Physical Sciences, Division of Mathematical Science (MPS/DMS)

- Nandini Kannan
- Christopher Stark

Deadline for HDR Solicitations

SOLICITATION	PRELIMINARY PROPOSALS	LETTER OF INTENT	FULL PROPOSALS
NSF 19-543 IDEAS LABS	March 4, 2019	N/A	BY INVITATION ONLY
NSF 19-549 FRAMEWORKS	N/A	N/A	May 7, 2019
NSF 19-550 TRIPODS	N/A	March 25, 2019	April 24 - May 08, 2019

Questions?

HDR-DIRSE@nsf.gov

Q&A-1: What Are Deadlines for each HDR Solicitation?

SOLICITATION	PRELIMINARY PROPOSALS	LETTER OF INTENT	FULL PROPOSALS
NSF 19-543 IDEAS LABS	March 4, 2019	N/A	BY INVITATION ONLY
NSF 19-549 FRAMEWORKS	N/A	N/A	May 7, 2019
NSF 19-550 TRIPODS	N/A	March 25, 2019	April 24 - May 08, 2019

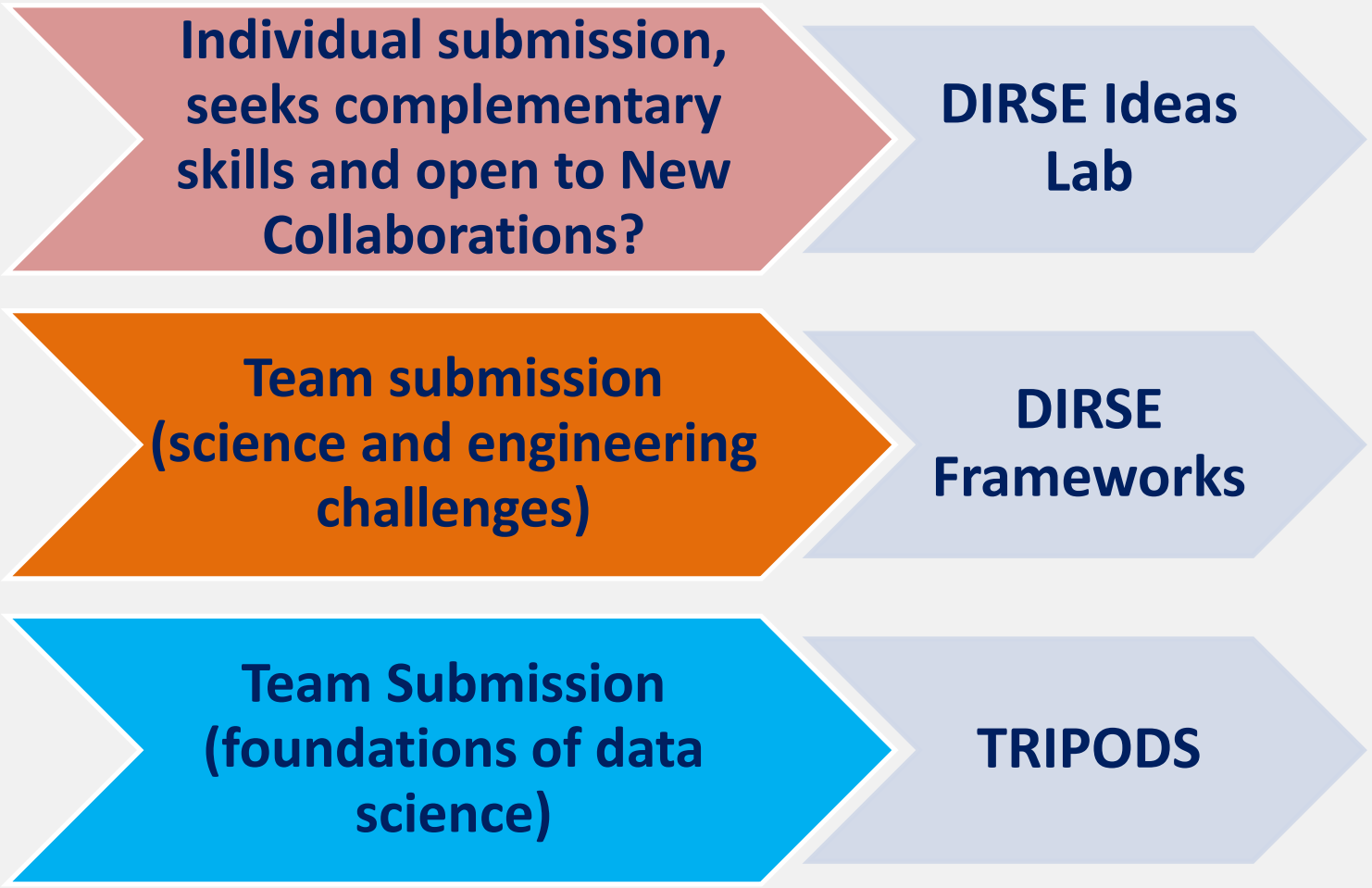
Q&A-2: Proposal Submittal

- When are proposals due?
  - Proposals must be received by **5 p.m. local time at the submitter's institution**.
  - Failure to submit by 5 p.m. submitter's local time will result in the proposal not being accepted.

- How do I submit a proposal to this program?
  - Please carefully read and follow the instructions provided in the solicitation itself (https://www.nsf.gov/pubs/2018/nsf18531/nsf18531.htm) and the NSF *Proposal & Award Policies & Procedures Guide (PAPPG)* available at (https://www.nsf.gov/pubs/policydocs/pappg18_1/index.jsp). If you need additional help preparing and submitting your proposal, we recommend that you contact your institution's Sponsored Projects Office.

- Do I need to use Grants.gov or Fastlane to apply?
  - You may use either Fastlane or Grants.gov for most solicitations;
  - Ideas Labs Preliminary proposals require FastlLane.

Q&A-3: Which Solicitation is Right for Me?



Q&A-4: Proposal Submittal Limits Across Solicitations

- If I am selected for the Ideas Lab, will I be able to submit a Frameworks proposal?
  - NO
- If I am selected for the Ideas Lab, will I be able to submit a TRIPODS proposal?
  - YES

Q&A: Ideas Lab

- If I am selected for the Ideas Lab, will I be able to submit a Frameworks proposal?
  - NO
- My collaborator and I have both applied to be part of the Ideas Lab. Will both of us be selected?
  - There is no guarantee that either or both of you will be selected

Q&A Frameworks: Proposal Limits

- If I am the PI on a proposal to Frameworks (NSF 19-549):
  - Can I be the PI on any other proposal to NSF 19-549? NO
  - Can I be a co-PI on any other proposal to NSF 19-549? NO
  - Can I be Senior Personnel on any other proposal to NSF 19-549? NO

- If I am a co-PI on a proposal to NSF 19-549 :
  - Can I be the PI on any other proposal to NSF 19-549? NO
  - Can I be a co-PI on any other proposal to NSF 19-549? NO
  - Can I be Senior Personnel on any other proposal to NSF 19-549? NO

- If I am Senior Personnel on a proposal to NSF 19-549 :
  - Can I be the PI on any other proposal to NSF NSF 19-549? NO
  - Can I be a co-PI on any other proposal to NSF 19-549? NO
  - Can I be Senior Personnel on any other proposal to NSF 19-549? NO

Q&A Frameworks: Proposal Limits

- If I am the PI on a proposal to Frameworks (NSF 19-549):
  - Can I be the PI on any other proposal to NSF 19-549? NO
  - Can I be a co-PI on any other proposal to NSF 19-549? NO
  - Can I be Senior Personnel on any other proposal to NSF 19-549? NO

- If I am a co-PI on a proposal to NSF 19-549 :
  - Can I be the PI on any other proposal to NSF 19-549? NO
  - Can I be a co-PI on any other proposal to NSF 19-549? NO
  - Can I be Senior Personnel on any other proposal to NSF 19-549? NO

- If I am Senior Personnel on a proposal to NSF 19-549 :
  - Can I be the PI on any other proposal to NSF NSF 19-549? NO
  - Can I be a co-PI on any other proposal to NSF 19-549? NO
  - Can I be Senior Personnel on any other proposal to NSF 19-549? NO

Q&A Frameworks: HDR and CSSI

Q: ***Two recent NSF solicitations -- HDR:I-DIRSE-FW (NSF 19-549) and CSSI (NSF 19-548) invite frameworks proposals. How are they different?***

A: While both solicitations involve researchers from diverse scientific backgrounds

- **HDR:I-DIRSE-FW (NSF 19-549)** targets science and engineering areas where an investment in data-intensive approaches has the maximum potential for a transformative impact.
- **CSSI (NSF 19-548)** seeks innovative cyberinfrastructure that will serve a research community. It emphasizes integrated cyberinfrastructure services, quantitative metrics with targets for delivery and usage of these services, and community creation.

A primary difference between the two is that HDR will conduct *new research* into data intensive approaches while CSSI will develop innovative mechanisms to deliver these approaches to the community in a robust way. An HDR Frameworks project may develop prototypes for testing their methods; a CSSI project is expected to develop a robust capability that serves the community.

Q&A TRIPODS

- 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 TRIPODS

- 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.

Backup Slides

Science and Engineering Challenges: Examples

- Near-term ecological forecasting;
- Understanding how the phenotype of living things is determined by their genotype and environment;
- Real-time sensing, learning, and decision making for resilient engineering systems;
- Development of autonomous technologies;

Science and Engineering Challenges: Examples

- Predictive understanding of the earth system which includes climate, weather, hydrologic, seismic, and space weather hazards;
- Multi-messenger astrophysics;
- Understanding the nature of dark matter;

Science and Engineering Challenges: Examples

- Predictive design of next-generation catalysts;
- Elucidation of design rules for emergent molecular properties from atomic-scale interactions;
- Design of sustainable chemical manufacturing systems;
- Real-time optimization and control of complex chemical and biological systems;
- Discovery of new advanced materials;

Science and Engineering Challenges: Examples

- Integration of heterogeneous data for explaining human behavior, learning, and social processes; understanding the brain, including prediction of complex systems for neuroimaging and neurological applications;
- Understanding student learning and success across STEM disciplines