# HDR$^2$

## From Harnessing the Data Revolution → To Harvesting the Data Revolution

NSF HDR PI Meeting

https://indico.cern.ch/event/1174814/

Zoom

Alexandria

Oct 26-27 2022

Welcome

# Community and Safety

- Be welcoming and open minded
- Communicate with empath
- Respect each other
- Recognize that intent does not equate to impact
- Strive to make informal activities relating to HDR inclusive to all
- Actively work to better our community

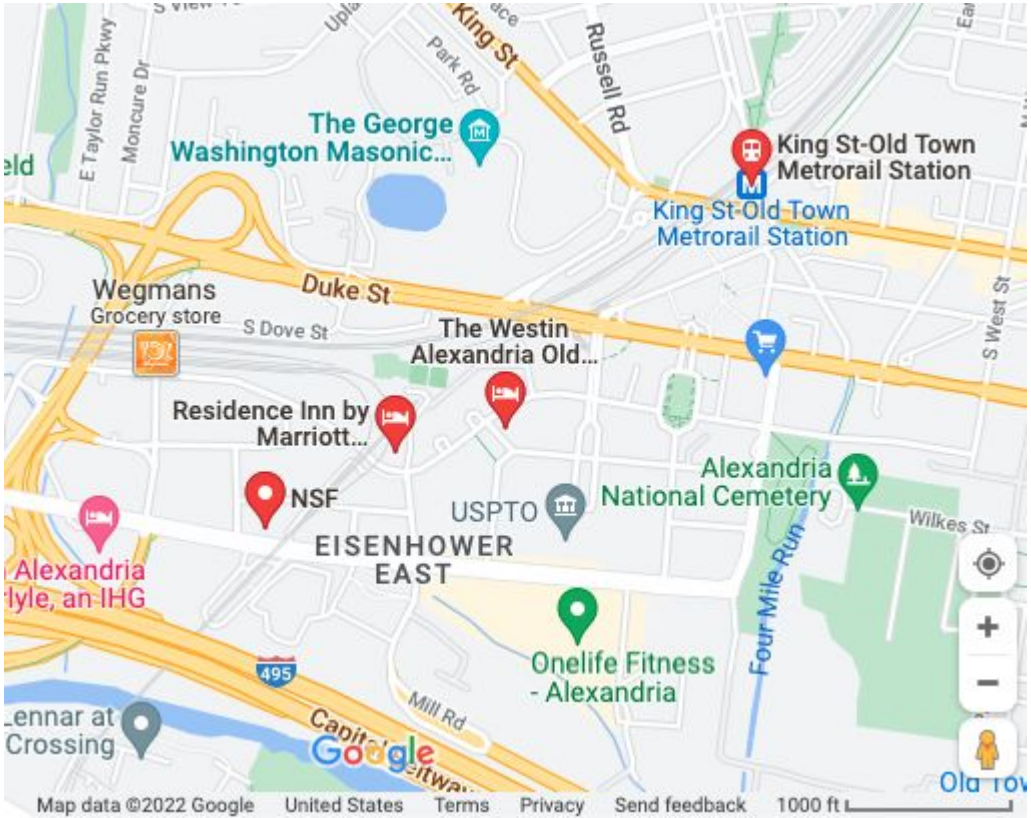
Code of Conduct

# Bringing us closer

# COVID-19 Prevention Plan


- Face coverings are strongly recommended
- Quarantine is no longer required after close contact with a person who has COVID-19.
- Self-monitoring for symptoms required daily (no reporting): Do not come to work or class, and follow instructions on the flowchart if you experience COVID-19 symptoms that are not attributable to another condition.
- Reporting a positive test result to the meeting organizer is required.
- Testing: Get tested if you develop symptoms

# Venue




All activities are in the
**Edison ABC** of
The Westin Alexandria Old Town

# Goal of the meeting

- How can we work together and eventually go from Harnessing the Data Revolution to Harvesting the Data Revolution or HDR²?
- To build a strong and sustainable HDR Ecosystem

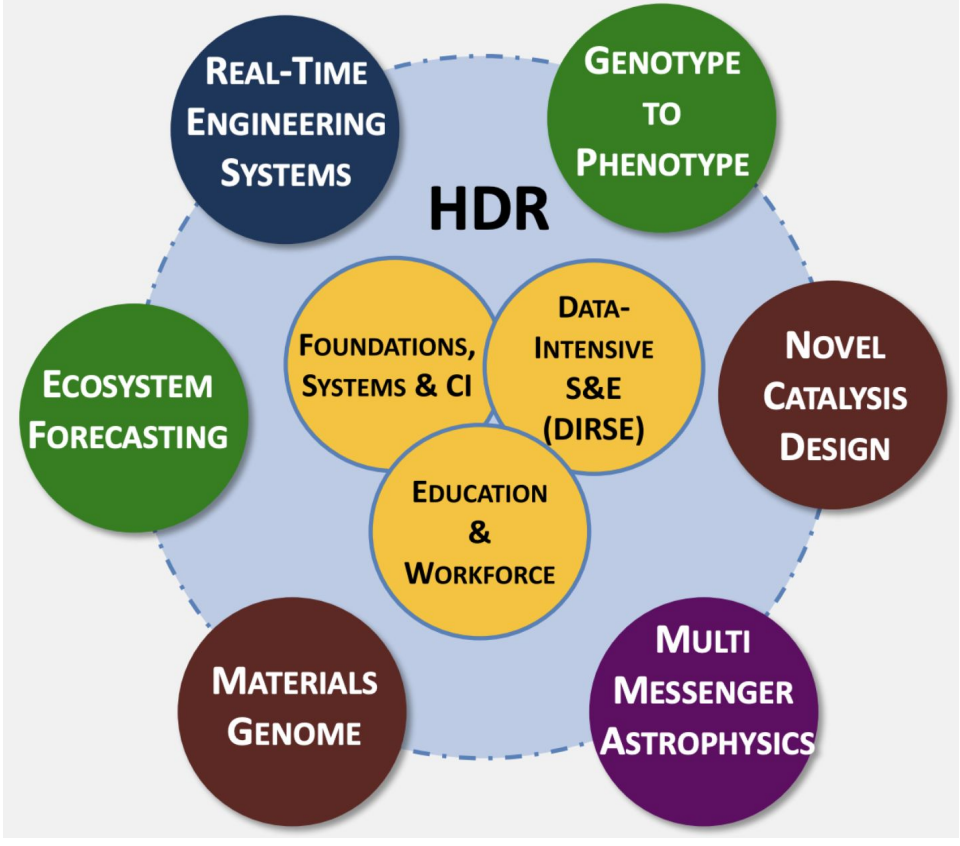
# Deliverables

- We would like to come up with ideas of community activities or proposals of work forces should be formulated in order to achieve the goal

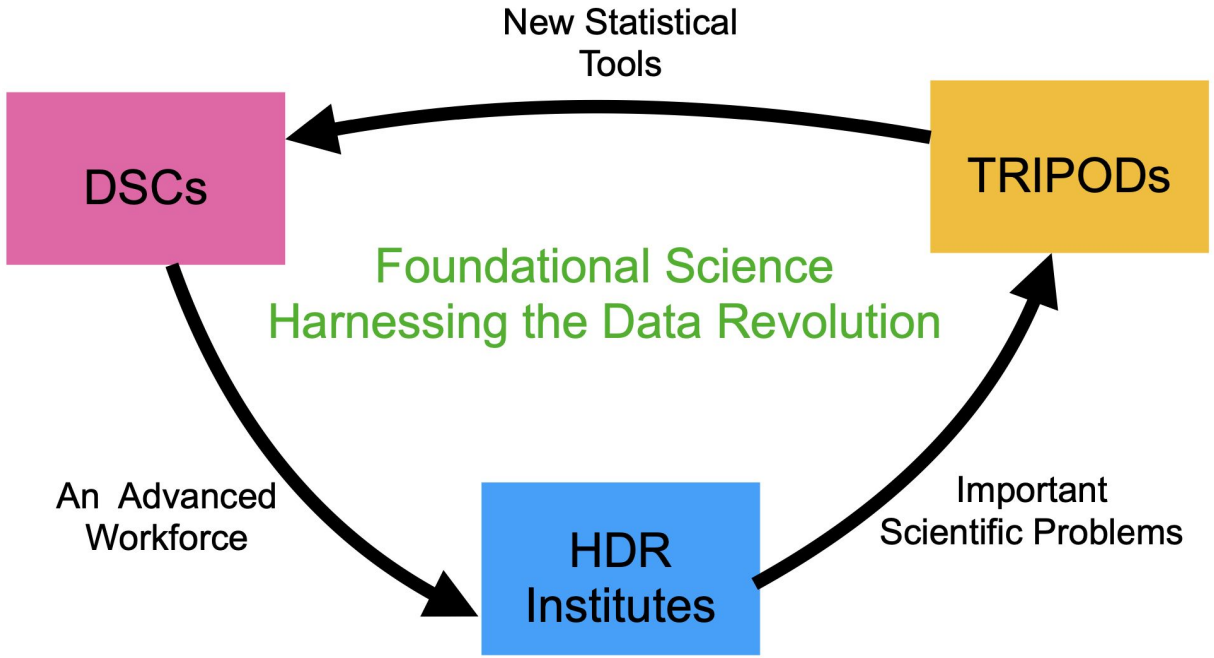
# Harnessing the Data Revolution (HDR)

Engaging 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



# NSF HDR Award Category




Data Science Corps

Transdisciplinary Research in Principles of Data Science

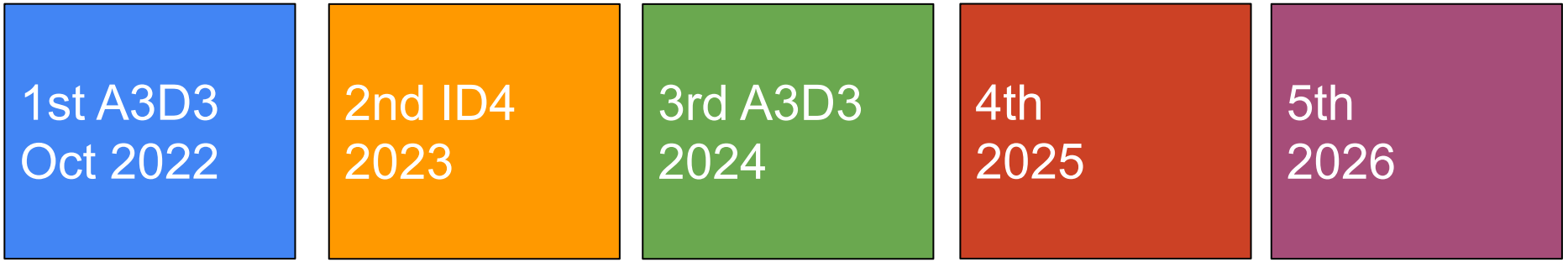
Institutes for Data-Intensive Research in Science and Engineering

# HDR Ecosystem supplement grant

- Supplemental funding ($1.25M) awarded to A3D3 toward efforts to grow and strengthen HDR Ecosystem
- Three Core initiatives in this supplement grant
  - HDR PI meetings
    - We are organizing this meeting, and 2024 edition
  - Machine Learning challenges
    - We are planning on initiating a series of ML challenges
  - Postbac enhancement
    - We would like to build on our existing program
    - Connect our work across the HDR community

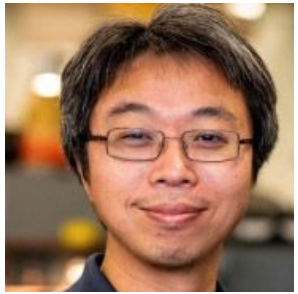
# NSF HDR PI meeting series

- 2022 Start from a core meeting within HDR community to formulate consensus of the HDR Ecosystem




- 2023 HDR PI meeting organized by ID4 to review progress
- 2024 Big HDR Conference Meeting UIUC

# Organization Committee for the First PI Meeting



Shih-Chieh Hsu, UW
(A3D3, co-chair)

Phil Harris, MIT
(A3D3, co-chair)

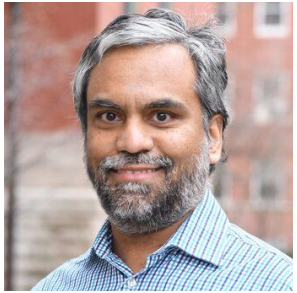
Valerie Varr,
Bard C.(DSC)

Nathan A. Quarderer,
CU Boulder (DSC)

Ambuj Singh,
UCSB (DSC)

Jane Greenberg,
Drexel U. (ID4)

Paula Mabee,
NEON (Imgeomics)



Anand Padmanabhan,
UIUC (I-GUIDE)

Jianwu Wang,
UMBC (iHARP)

Barna Saha, UCSD
(TRIPODS)

# NSF Advisory Committee

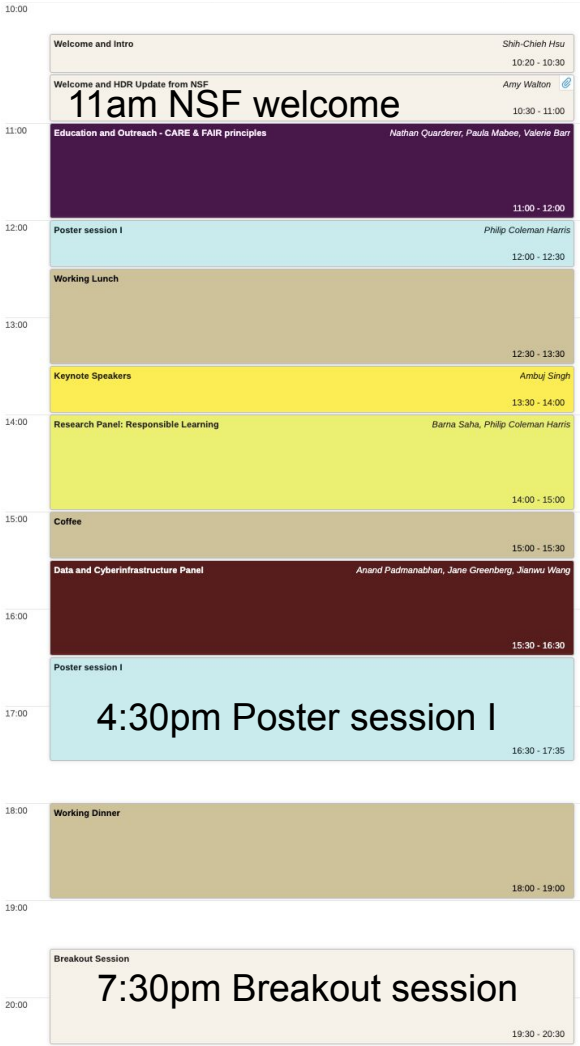
Amy Walton

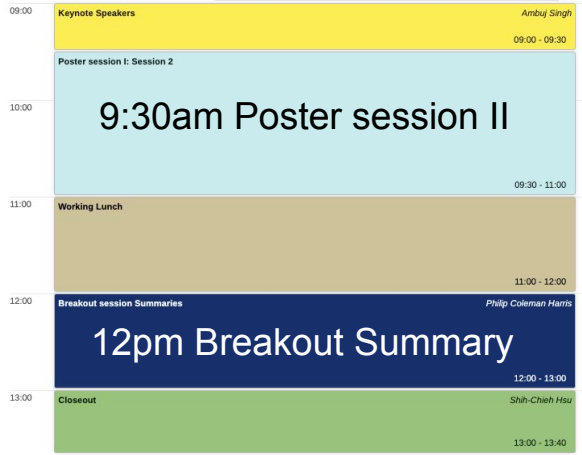
Vyacheslav Lukin

Sylvia J. Spengler

# Agenda

- Two Keynote talks: Juliana Freire and Sarah Stone



# Activities in this PI meeting

- Panel discussions on three principle components:
  - Research
  - Education & Outreach
  - Data & Cyberinfrastructure
- Lightning talk
  - 1 min for DSC or TRIPODS Phase I
  - 2 min for Institutes or TRIPODS Phase II
- Poster sessions
  - This is a great opportunity to know more about each representative and their efforts
- Breakout sessions
  - Propose activities or formulation of working groups
  - Each proposal should include a coordination plan (candidate coordinators, milestone, timeline, anticipated deliverables).
  - Progress on each activity will be reported in the future PI meeting

# Breakout sign-up

- Sign-up sheet will be open during afternoon poster session
- Policy of member compositions:
  - at least one member from each award category
  - maximum number of people is 8
- Results will be summarized in the Day 2 plenary session

[Link]

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| | Research1: Inductive bias | Research2: Complex data structure | Research3: Machine Learning challenge | Data & Cyberinfrastructure1: Data incompleteness, missing data | Data & Cyberinfrastructure2: Computational Reproducibility | Data & Cyberinfrastructure3: Data and code scalability | Education & Outreach1: Pedagogy in teaching data science & ethics | Education & Outreach2: Reproducible research and sustainable resource for sharing code and data in training program | TBA |
| #1 Lead | Jeremias Sulam | Vandana Janeja | Phil Harris | Eric Torber | Anand Padmanabhan | Jianwu Wang | Valeria Barr | Nathan Quarderer | |
| #2 DSC | | | | | | | | | |
| #3 TRIPODS | | | | | | | | | |
| #4 Institute | | | | | | | | | |
| #5 | | | | | | | | | |
| #6 | | | | | | | | | |
| #7 | | | | | | | | | |
| #8 | | | | | | | | | |

*Enjoy the meeting!*