On the Curation of Artifacts in the Era of AI/ML for Cybersecurity

ACSAC 2022

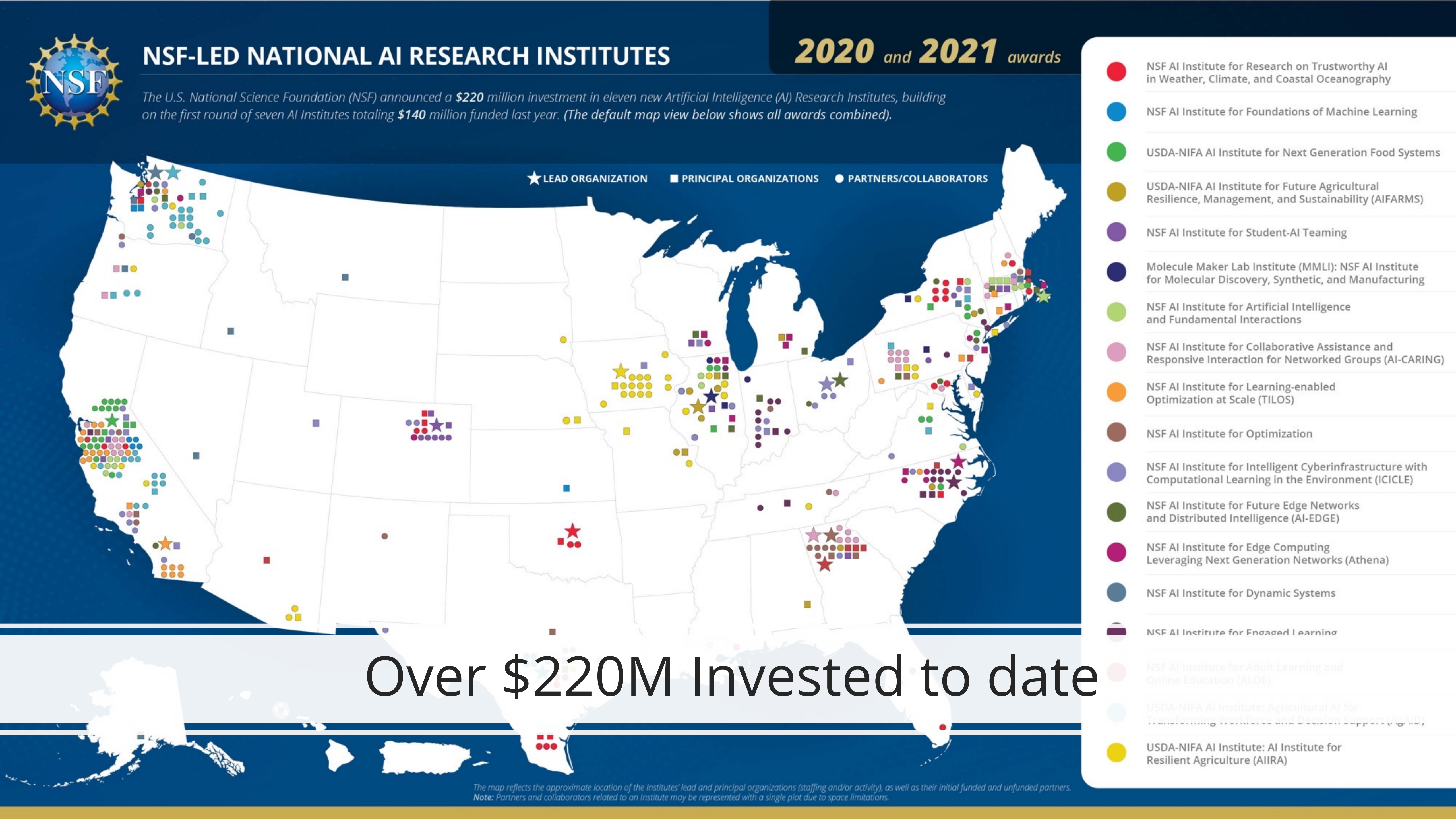
December 8, 2022

Robert Beverly

Program Officer, Office of Advanced Cyberinfrastructure

Position Statement

- Unique characteristics of cybersec make data for AI/ML challenging
  - Adversarial model, rare events, false positive rates, etc.
  - (see: Sommer and Paxson, "*Outside the Closed World*", IEEE S&P 2010
- There remains a lack of canonical datasets
  - Akin to e.g., MNIST
  - Hampers research
- Equitable access to data and artifacts missing
  - Data fiefdoms
  - Impacts science and research
  - Impacts academic careers



NSF-LED NATIONAL AI RESEARCH INSTITUTES

2020 *and* **2021** *awards*

*The U.S. National Science Foundation (NSF) announced a **$220** million investment in eleven new Artificial Intelligence (AI) Research Institutes, building on the first round of seven AI Institutes totaling **$140** million funded last year. (The default map view below shows all awards combined).*

- ● NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography
- ● NSF AI Institute for Foundations of Machine Learning
- ● USDA-NIFA AI Institute for Next Generation Food Systems
- ● USDA-NIFA AI Institute for Future Agricultural Resilience, Management, and Sustainability (AIFARMS)
- ● NSF AI Institute for Student-AI Teaming
- ● Molecule Maker Lab Institute (MMLI): NSF AI Institute for Molecular Discovery, Synthetic, and Manufacturing
- ● NSF AI Institute for Artificial Intelligence and Fundamental Interactions
- ● NSF AI Institute for Collaborative Assistance and Responsive Interaction for Networked Groups (AI-CARING)
- ● NSF AI Institute for Learning-enabled Optimization at Scale (TILOS)
- ● NSF AI Institute for Optimization
- ● NSF AI Institute for Intelligent Cyberinfrastructure with Computational Learning in the Environment (ICICLE)
- ● NSF AI Institute for Future Edge Networks and Distributed Intelligence (AI-EDGE)
- ● NSF AI Institute for Edge Computing Leveraging Next Generation Networks (Athena)
- ● NSF AI Institute for Dynamic Systems
- ● NSF AI Institute for Engaged Learning
- ● NSF AI Institute for Adult Learning and Online Education (ALOE)
- ● USDA-NIFA AI Institute: Agricultural AI for Transforming Workforce and Decision Support (AgAID)
- ● USDA-NIFA AI Institute: AI Institute for Resilient Agriculture (AIIRA)

Over $220M Invested to date

The map reflects the approximate location of the Institutes' lead and principal organizations (staffing and/or activity), as well as their initial funded and unfunded partners.
***Note:** Partners and collaborators related to an Institute may be represented with a single plot due to space limitations.*

NSF 23-517: Cybersecurity Innovation for Cyberinfrastructure (CICI)

- **Reference Scientific Security Datasets (RSSD**): *Projects in this program area should leverage instrumented cyberinfrastructure to capture metadata from scientific workflows and workloads as reference data artifacts that can help support reproducible security research, testing and evaluation*.
- **Example RSSD projects**:
  - "LaSIC: Labeled Security Information Capture", 2232864/Papadopoulos
  - "Massive Internal System Traffic Research Analysis and Logging", 2232819/Biever
- **Deadline**: February 17, 2023

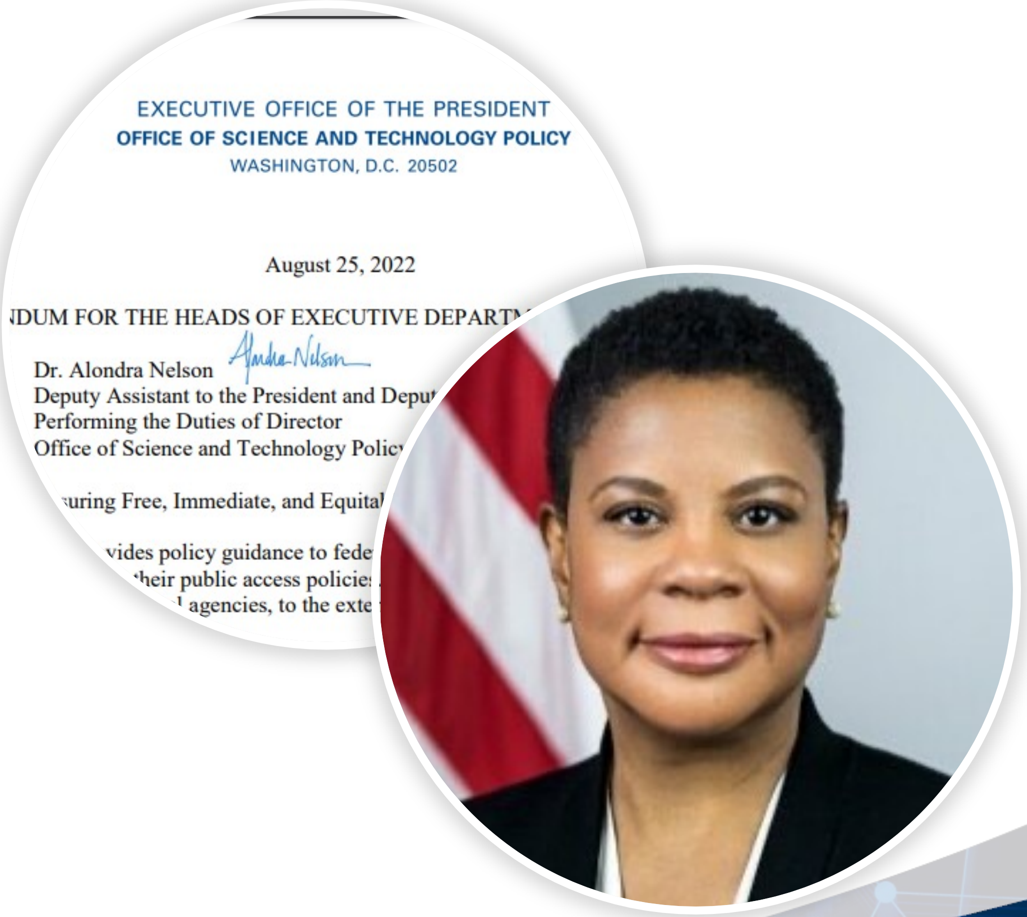
2022: White House Office of Science and Technology (OSTP) guidance on Open Science and Public Access, commonly called the “Nelson” Memo

Issued by OSTP
Acting Director Alondra Nelson

- calls for **Free**, **Immediate**, and **Equitable** public access

- new Public Access/Open Science plan by Feb 2023 (with policies by 2024, and implementation by 2025)

- default **zero-embargo** of peer-reviewed articles and underlying data



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

NDUM FOR THE HEADS OF EXECUTIVE DEPARTM

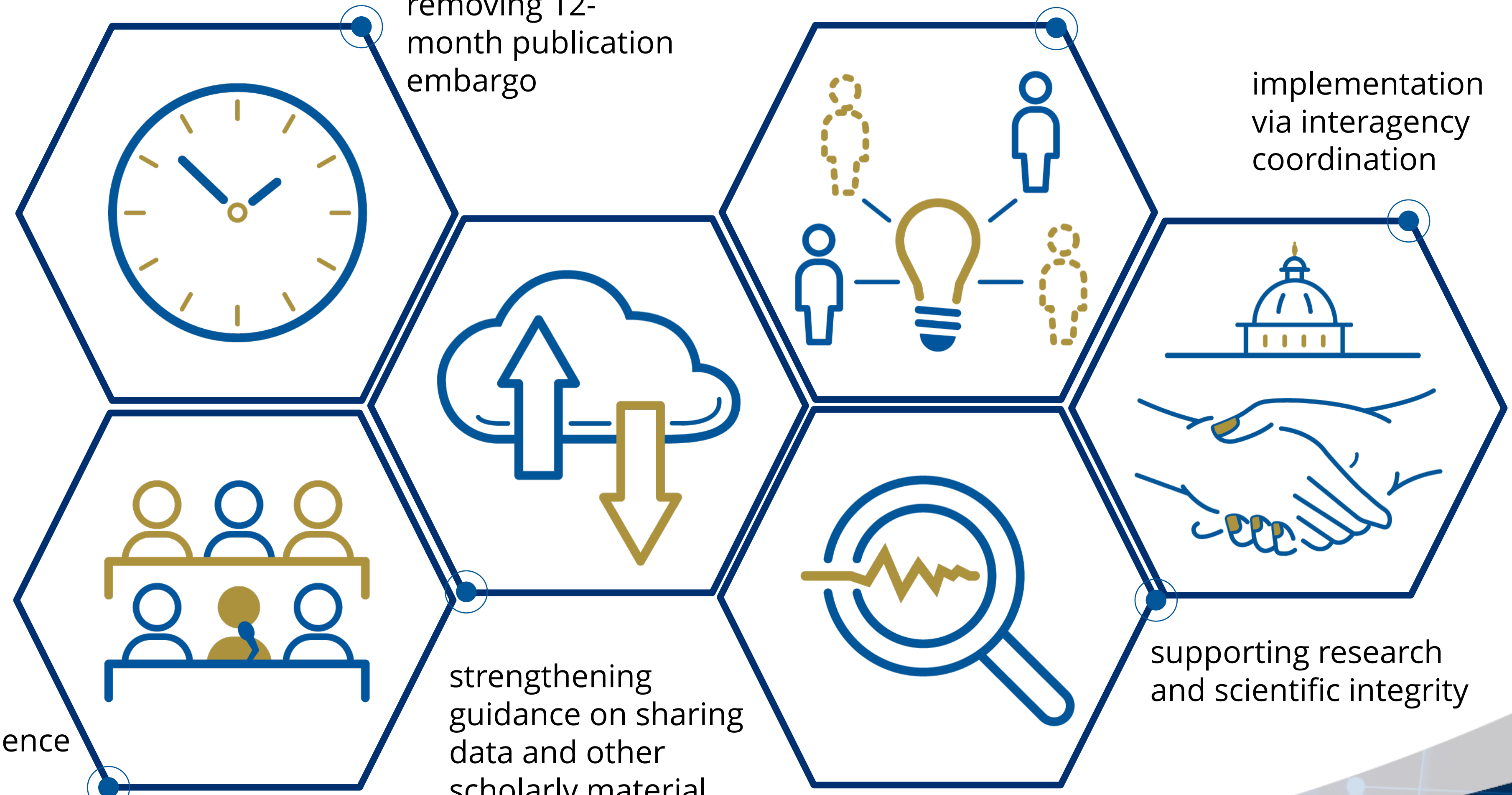
Dr. Alondra Nelson Alondra Nelson
Deputy Assistant to the President and Deput
Performing the Duties of Director
Office of Science and Technology Polic

uring Free, Immediate, and Equital

vides policy guidance to fede
heir public access policie
agencies, to the exte

What is the new policy guidance?

> *"American investment in such research is essential to the health, economic prosperity, and well-being of the Nation. There should be no delay between taxpayers and the returns on their investments in research."*
>
> *~ Dr. Alondra Nelson*



removing 12-month publication embargo

providing equitable access

implementation via interagency coordination

supporting research and scientific integrity

strengthening guidance on sharing data and other scholarly material

informed by years of OSTP Subcommittee on Open Science Leadership (currently led by OSTP, NASA, NIH, and NSF)