What is the value of Open Data / Software Sharing in a world of Open Science? The Support System – a view from NSF

***Allen Pope**, Program Director*
NSF Directorate for Geosciences, Office of Polar Programs

April 19, 2023

AGU/CHORUS FORUM: How open are data & software?

Growing importance of open, accessible, and reproducible science at NSF

- **OSTP Memo**: *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research* (Aug. 25, 2022)

- **FAIROS RCNs** (NSF 22-553)
  - New NSF program supporting Research Coordination Networks (RCNs) that advance FAIR principles (findable, accessible, interoperable, reusable) and open science (OS) practices

- **Reproducibility & Replicability DCL** (NSF 23-018)

Federal Year of Open Science
https://open.science.gov/

NSF 23-018

Dear Colleague Letter: Reproducibility and Replicability in Science

October 25, 2022

Dear Colleagues:

A 2019 consensus study report published by the National Academies of Sciences, Engineering, and Medicine (NASEM) discussed the meaning of the terms replicability and reproducibility and identified approaches for researchers, academic institutions, journals, and funders to improve reproducibility and replicability in science [1]. In July 2021, at NSF's request, NASEM convened an expert meeting focused on National Science Foundation (NSF) policies and investments to make reproducible and replicable science easier for scientific communities to understand and execute and to embed reproducibility and replicability within the fundamental scientific method.

Through this Dear Colleague Letter (DCL), NSF reaffirms its commitment to advancing reproducibility and replicability in science. NSF is particularly interested in proposals addressing one or more of the following topics:
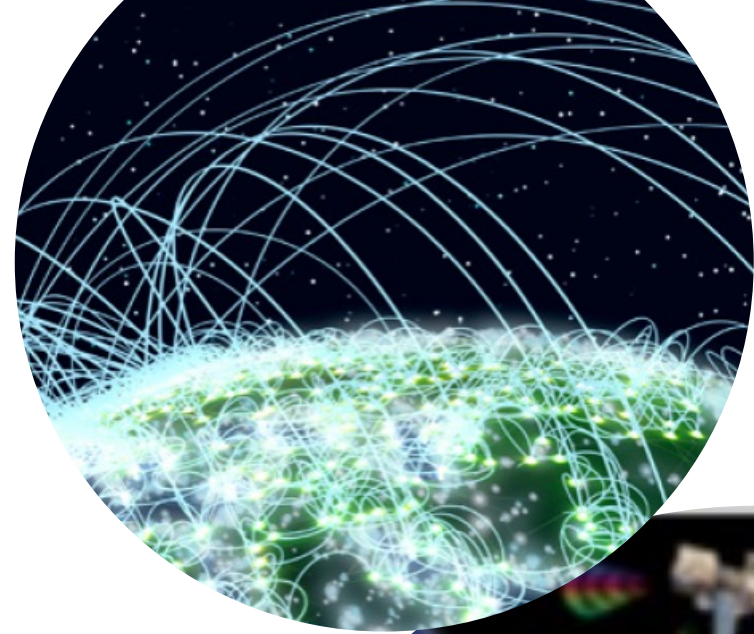
1. **Advancing the science of reproducibility and replicability.** Understanding current practices around reproducibility and replicability, including ways to measure reproducibility and replicability, what reproduction and replication means in practice, the right degree of replicability to target, quantitative measures of progress to understand the effectiveness of interventions to improve reproducibility and replicability, and exploration of reasons why studies may fail to replicate.
2. **Research infrastructure for reproducibility and replicability.** Developing and facilitating adoption of cyberinfrastructure tools and/or research methods that enable use of reproducible and replicable practices across one or more science and engineering communities.
3. **Educational efforts to build a scientific culture that supports reproducibility and replicability.** Enabling training in science and engineering communities to identify and encourage best practices for reproducibility and replicability, providing community-building and institutional support, and supporting broad public outreach about rigor, reproducibility, and replicability in science.

Investigators who wish to submit proposals on any of these topics, or others related to advancing reproducibility and replicability in research, are encouraged to reach out to programs and program officers to discuss the fit of their ideas to existing funding opportunities. Definitions of the terms replicability and reproducibility may be found in Reference [1].

GEO Cyberinfrastructure Goals

- **Advance geoscience research**
- **Promote openness and participation through Open Science**
- **Pursue AI/ML innovation in the geosciences**
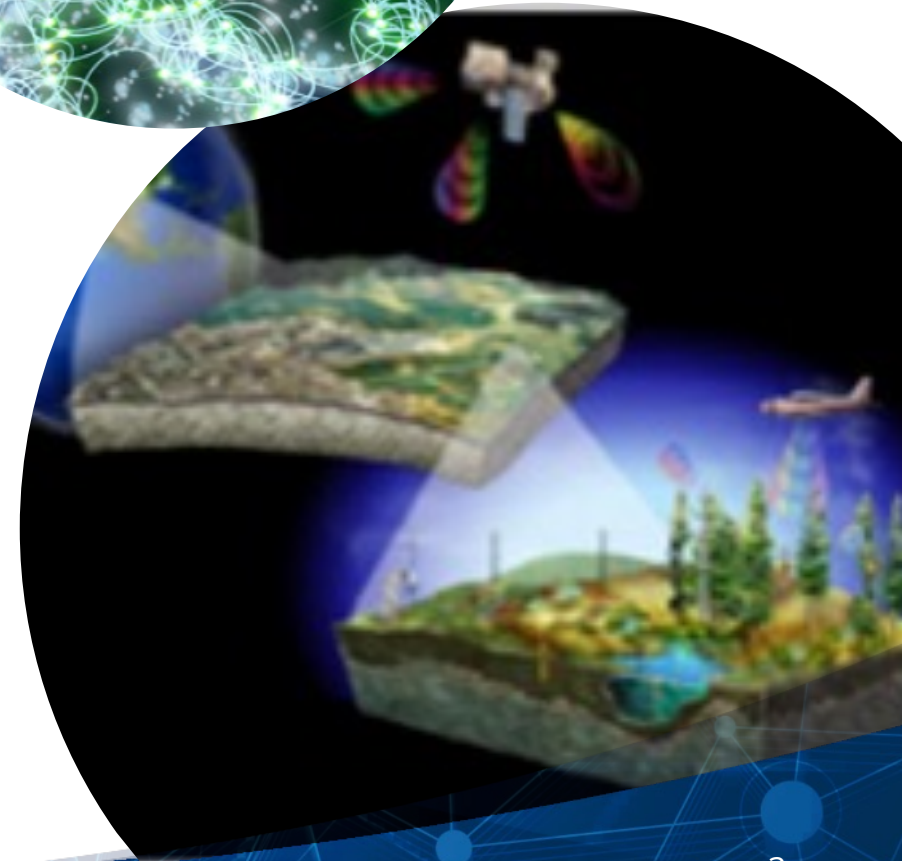
NSF's FY 2023 Budget Request:
Major GEO Investments

NSF-Wide Climate Activities: $33.67M

GEO Climate Incubators: $80M

GEO Cyberinfrastructure Incubator: $20M

GEO Education, Diversity, and Equity Incubator: $7.5M

Foundational Investments

Office of Advanced Cyberinfrastructure (OAC) opportunities

- **Cyberinfrastructure for Sustained Scientific Innovation (CSSI)** (NSF 22-632): Support for robust, reliable, and sustainable data and software cyberinfrastructure (*Deadline:* December 16, 2022)

- **CyberTraining (NSF 23-520):** Supports efforts toward broad adoption of CI tools, methods, and resources; and integration of CI literacy into curriculum / instructional materials (*Deadline:* Feb. 23, 2023)

- **SCIPE (NSF 23-521):** "Strengthening the Cyberinfrastructure Professionals (CIP) Ecosystem" (*Deadline:* Feb. 23, 2023)

- *OAC CI Resource Ecosystem:*
  - **ACCESS**: Coordinates allocations for advanced computing, visualization, and data resources for researchers and educators (follow-on to XSEDE) (https://access-ci.org)
  - **CloudBank:** Enabling access to commercial cloud service resources
  - **Partnership to Advance Throughput Computing (PATh):** Pilot support for high-throughput computing (HTC) resources (see NSF 22-051)

More information here: https://www.nsf.gov/geo/geo-ci

Relevant CI-Related Programs

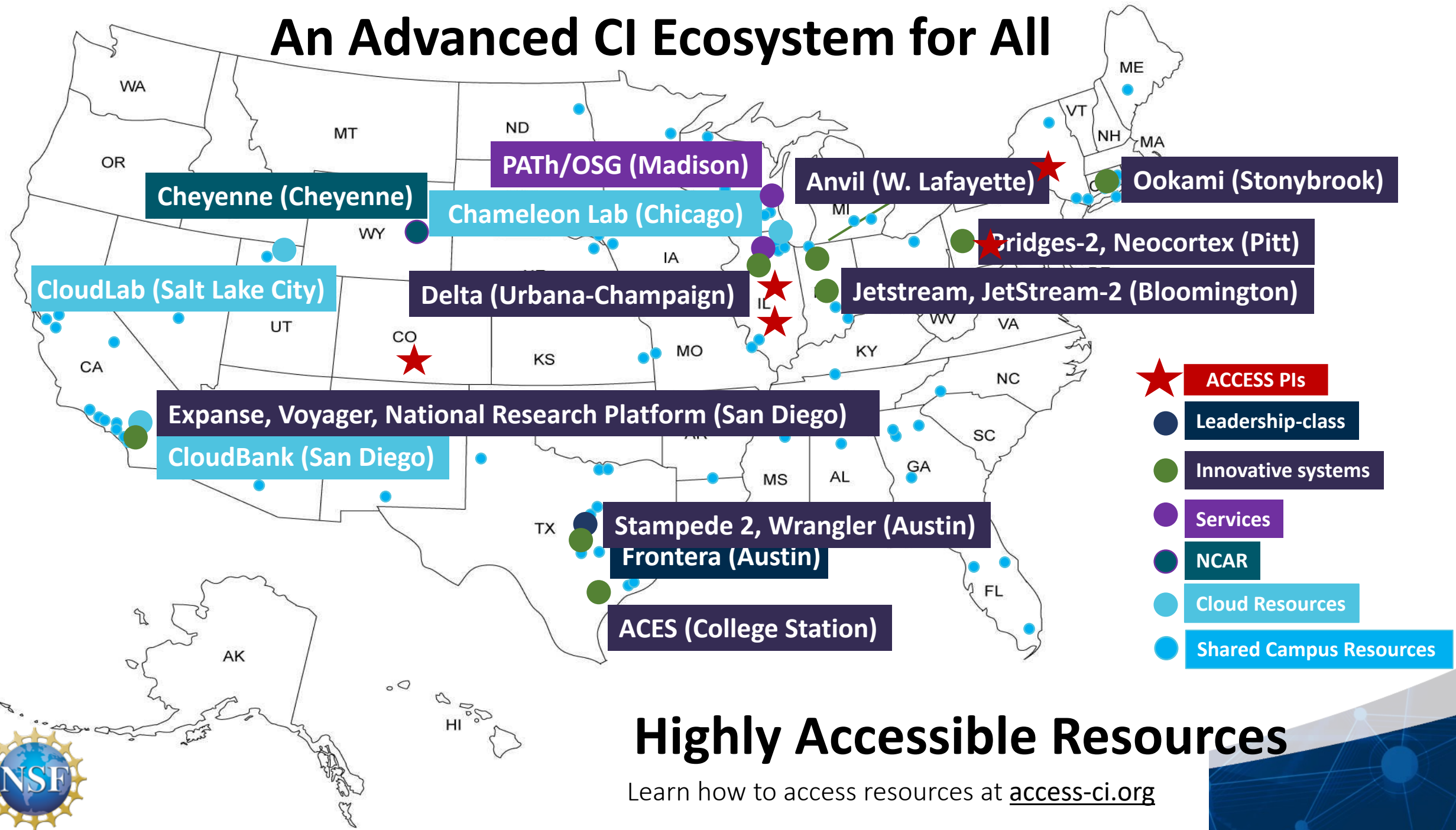|	CSSI	OAC CORE	SCIPE	CyberTraining
What will your project focus on?	Develop software or data repository	Perform research that will enable future CI	Support research that uses CI, foster CIP careers	Provide training
What gap will you fill?	Increase community CI for research or education	Increase knowledge needed for CI	Increase support for CIP career & research by underserved groups	Increase research workforce to leverage CI
What will your project deliver?	Community-sustained CI	Techniques that will enable CI and a CI prototype	Research support in CI, CIP career paths	Scalable and sustainable training program

Programs have specific purposes, however, are not necessarily mutually exclusive

An Advanced CI Ecosystem for All

PATh/OSG (Madison)
Anvil (W. Lafayette)
Ookami (Stonybrook)
Cheyenne (Cheyenne)
Chameleon Lab (Chicago)
Bridges-2, Neocortex (Pitt)
CloudLab (Salt Lake City)
Delta (Urbana-Champaign)
Jetstream, JetStream-2 (Bloomington)
Expanse, Voyager, National Research Platform (San Diego)
CloudBank (San Diego)
Stampede 2, Wrangler (Austin)
Frontera (Austin)
ACES (College Station)

ACCESS PIs
Leadership-class
Innovative systems
Services
NCAR
Cloud Resources
Shared Campus Resources

Highly Accessible Resources

Learn how to access resources at access-ci.org

Technology, Innovation, and Partnerships (TIP):
CI-related programs

TIP is NSF's newest directorate. Supports use-inspired research, accelerating development of key technologies, and expanding the STEM workforce

- **Convergence Accelerator:** phased funding model to accelerate solutions toward societal impact (RFI for topics => Workshops => topic selection for project cohorts)
  - *2022 workshop examples:* Ethical Design of AI, Computing Solutions for Climate-Driven Extreme Events
  - *Past cohort examples:* Open Knowledge Networks, AI-driven Innovation via Data & Model Sharing

- **Innovation Corps (I-Corps™)** (NSF 21-552)
  - Up to $50k for intense 7-week training in customer discovery
  - Anyone with active NSF award within last 5 years may apply

- **Pathways to Enable Open-Source Ecosystems (POSE)** (NSF 22-552)
  - Supports efforts toward harnessing open-source development approaches for new technology solutions to problems of national and societal importance
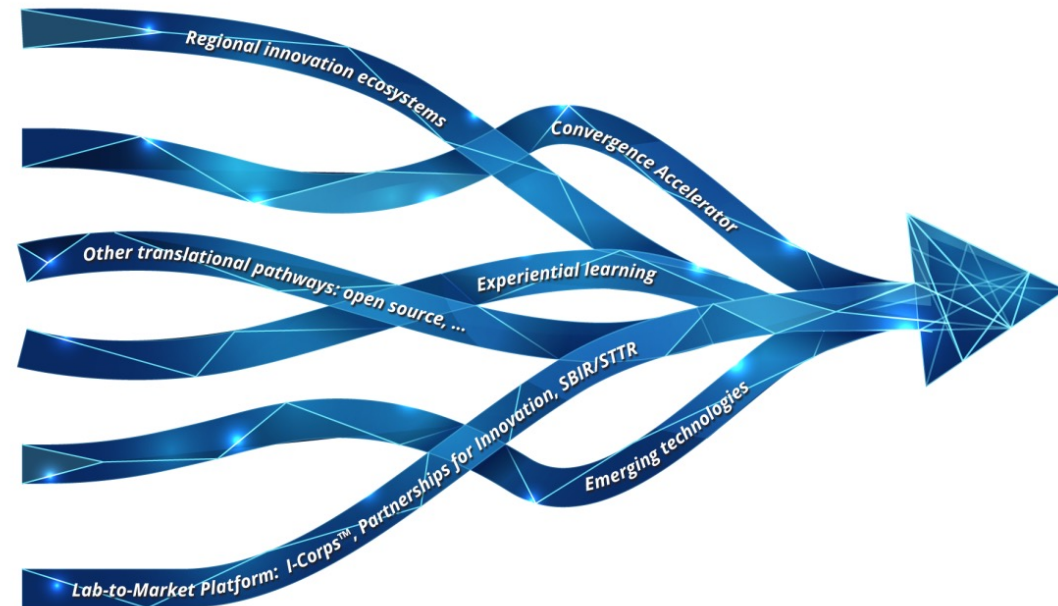
Directorate for Geosciences (GEO):
Support for cyberinfrastructure (CI)

GEO-wide:

- **GEO Open Science Ecosystem (OSE)** Supports the improvement and democratization of access to open-science resources in the geosciences — with a focus on increasing access to data, physical collections, software, advanced computing and other resources. (NSF 23-534)

Within GEO Divisions:

- In **Earth Sciences (EAR),** the **Geoinformatics program** (NSF 21-583) supports widely-used CI services relevant to Earth Scientists

- In **Polar Programs (OPP)**, Polar CI proposals may be submitted through **Antarctic Research** (NSF 21-567) or **Arctic Research Opportunities** (NSF 21-526)

- In **Atmospheric & Geospace Sciences (AGS)** and **Ocean Sciences (OCE)**, development of CI and cyber tools are supported through core scientific and facilities programs

More information here: https://www.nsf.gov/geo/geo-ci

Supporting Open Polar Research Software (23-053)

- Invites the sustainable development and use of open source software, tools, libraries and frameworks that are critical for polar scientific objectives.

- Builds upon recently updated data policy which includes code

- Leverages open data & increasing accessibility/inclusivity of polar science

- Upgrades, translate, document, build community, train & much more

📷: Amanda Kowalski

A couple parting thoughts / questions…

- Open is not a binary, and researchers are challenged in getting more open.
  - Open data → Open software
- Cyberinfrastructure is people, too!
- If you build it & they will come is no longer good enough for CI
  - Adaptable workflows that can be modified & build on previous work
- What does sustainability look like? We can only fund what comes in!!