2024 NSF CSSI-Cybertraining-SCIPE PI Meeting
August 12 to 13, 2024, Charlotte, NC

Link to meeting: https://confmeet.github.io/ccs2024/

Co-Chairs:
Abani Patra (Tufts University), Mary Thomas (University of California San Diego)

Program Committee Members:
Elias Bou-Harb (Louisiana State University, CSSI), Jeffrey Carver (University of Alabama Tuscaloosa, CSSI), Yuebin Guo (Rutgers University, CSSI), Ratnesh Kumar (Iowa State University, CSSI), Julien Langou (University of Colorado Denver, CSSI), Guoyu Lu (SUNY Binghamton, CSSI), Vivak Patel (University of Wisconsin, CSSI), Marianna Safronova (University of Delaware, CSSI), Isla Simpson (NSF NCAR, CSSI), Dhruva Chakravorty (Texas A&M, CT), Jane Combs (University of Cincinnati, CT), Hantao Cui (Oklahoma State, CT), Sushil Prasad (UT San Antonio, CT), Adnan Rajib (UT Arlington, CT), Susan Rathbun (San Diego Super Computer Center, CT), Erik Saule (University of North Carolina – Charlotte, CT), Isla Simpson (UCAR, CT), Alan Sussman (University of Maryland, CT), Shaowen Wang (University of Illinois Urbana-Champaign, CT), Sarina (Zhe) Zhang (Texas A&M, CT)

Additional Contributors to Sessions and Report Preparation:
Katie Antypas (NSF), Ritu Arora (Wayne State), Ben Brown (DOE/OASCR), Varun Chandola (NSF), Daniel Crawford (Virgina Tech), Ian Foster (U. Chicago/ Argonne), Dave Hart (NCAR), Mike Heroux (Sandia), Mary Ann Leung (Sustainable Horizons Institute), Benjamin Lynch (U Minn), Dan Negrut (UW Madison), D. K. Panda (OSU), Manish Parashar (Utah), Melissa Kline Struhl (MIT), George K. Thiruvathukal (Loyola University Chicago)

Abstract The second annual NSF/OAC CSSI/CyberTraining and related programs PI meeting was held August 12–13 in Charlotte, NC, with participation from PIs or representatives of all major awards. Keynotes, panels, breakouts, and poster sessions allowed PIs to engage with each other, NSF staff, and invited experts. The 286 attendees represented 292 awards across CSSI, CyberTraining, OAC Core, CIP, SCIPE CDS&E, and related programs, and presented over 250 posters. This report documents the meeting's structure, findings, and recommendations, offering a snapshot of current community perspectives on cyberinfrastructure (CI). A key takeaway is a vibrant, engaged community advancing science through CI. AI-driven research modalities complement established HPC and data-centric tools. Workforce development efforts align well with the CSSI community.

Disclaimer. This report is based on activities supported by the National Science Foundation under award numbers 2435580 and 2434556. Any opinions, findings, conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements We acknowledge the efforts of the following without whose contributions above and beyond the call of duty this meeting would simply not have happened!

- Our NSF program directors who supported us with their time and resources.
- Susan Rathbun, Jordan Wilkinson and Meghan Rodriguez whose efforts were crucial to the organization and smooth running of the meeting.

Contents

1 Executive Summary

1.1 Summary

The NSF/OAC CSSI/Cybertraining and related programs PI meeting was successfully conducted August 12-13 in Charlotte, NC[1]. The meeting was attended by PIs and/or representatives of all major awards from both of these programs. This was the second joint CSSI and CyberTraining meeting [2]. Keynote, panels, breakouts and required posters provided PIs opportunites to learn from each other, NSF personnel and distinguished invitees. The purpose of this report is to document meeting processes and record findings and recommendations that provide a snapshot of the community thinking about these programs and cyberinfrastructure current state and futures.

The primary finding from the different activities at the meeting and inputs provided is a thriving and engaged community of scientists at the intersection of cyberinfrastructure(CI) and science building, sustaining and CI for science. Newer research modalities driven by AI complement the robust HPC and data driven CI enabled scientific tools. Training and workforce development communities integrate well with the CSSI community.

1.2 Future Directions

Overall Future Directions Attending PIs used part of the breakout times to define desired future directions. In a desired future there will be a skilled and well trained workforce capable of enabling and sustaining the ML/AI transformation, supported by low barrier access to tools, data, flexible and scalable storage solutions, science workflows and high-performance computing resources through advanced CI. This will enable several high priority scientific use cases in fields requiring large-scale data processing and long-term storage like environmental modeling, genomic research, and AI-driven simulations in physics (secs. 5.1.2, 5.2.3 and 5.4.3 for details). Beyond these domains the use of AI in science is still in its infancy and good CI can enable use of AI in many fields (sec. 5.4.3). Effective translation of science advances to tools requires a complete ecosystem(hardware, tools and people) and culture enabling attributes like reproducibility in science done with computing (sec. 5.7.3). Successful CI will therefore have incentives for developers supported write better scientific software across domains (sec. 5.9.3). While, the programs in this cohort are developing distinct communities that focused on critical albeit different aspects of computationally enabled research integration and strengthening structures within and across communities can enable better science (secs. 5.10.3 and 5.11.5). Finally, a truly desired outcome is the recognition and formalization of the roles of Research Software engineers as co-equal to research staff, closely integrated and integral to making the science happen - through both career recognition, citations, as well as stable career paths (sec. 5.3.3).

1.3 Recommendations

Overall Recommendations

R1 It is recommended that the organizing committee is appointed early and meeting planning and communication to attendees start 6-9 months before meeting time. Cost (registration and travel) and convenience of attending (travel, hotel and dates) are high priorities for attendees. Resources from small registration fee were crucial to hosting meeting seamlessly with many unanticipated costs or costs that are difficult to charge to NSF grants.

R2 Poster sessions are well received and provide meaningful opportunities for networking and sharing of best practices among diverse groups of researchers from many areas and backgrounds and NSF personnel. Therefore, the experience at posters should be high priority. Breakouts are effective means of generating community input with good feedback on diverse topics but good breakouts need structure and time. Panels are less effective at holding PI's attention though considerable expenditure is entailed in getting good panelists.

R3 Investments leading to low barrier access to high end CI, innovative approaches to provisioning training on ML/AI methods, supporting standardization and interoperability and long term engagement of necessary personnel are needed for enabling the ML/AI transformation of science. Furthermore, there needs to be clear domain-specific standards for data collection, data curation and data sharing such that the resulting data can be used to train AI including foundation models for use integral to advanced CI. Significant investment is required to set up and operate such data frameworks.

R4 Small EAGER like grants for hardware maintenance, expanded training in supporting heterogeneous computing resources, funding for improving AI hardware effectiveness and encouragement for local, regional and national "condo" models for computing hardware more accessible for the sciences.

R5 The critical role of RSEs in the CI ecosystem needs to be carefully supported through diverse funding models, treating RSE support as integral to major CI investments. Matching of RSEs to domain sciences builds expertise and beneficial relationships.

R6 Integrating CI specialists and domain science experts in training opportunities can break silos. Including graduate students and early career scientists in such training is desirable.

R7 To support translation of science advances to tools, encourage mechanisms for providing credit for software and tools like DOIs for software. Encourage other directorates to value science tools even if supported by OAC division. Provide sustainability grants for science and emphasize a culture of reproducibility. It is also recommended that a clearer path for integrating CSSI software into ACCESS and future NSF supported resources be created.

R8 Community building requires refined and expanded metrics developed in collaboration with social scientists on both quantitative and qualitative aspects of community engagement for better measurement of community sentiment and long term engagement.

R9 Advanced CI needs community-driven and innovative curricula to adapt to rapidly changing technology with multiple delivery formats from reskilling like the Software Carpentries to diverse targets ranging from K-12, community colleges and current CI professional from academia and industry.

R10 Encourage inter-directorate and industry Collaboration with other NSF directorates, such as EDU and domain science directorates. In particular partnering with the NSF TIP directorate and industry partners to develop tools and training which is valuable to both science professionals and industry for staff upskilling.

Detailed Recommendations

- **Meeting Organization** and Running From "Notes for Next year's (2025) planning" document compiled as we organized and ran the meeting.
  - Start planning earlier! This year's meeting decision making was dominated by the deadlines.
  - List the NSF program affiliation on the name badge (CSSI, SCIPE, Cybertraining, etc)
  - Post web info in more places (add to slides)
  - Bell or some way to get people to leave poster session and go to next sessions
  - Poster session was very loud (good) - maybe leave more space between poster rows so conversation is easier.
  - CaRCC professionalization and workforce development presentation?
  - Short URLs for breakout sessions scribes
  - Full A/V in breakout rooms
  - Extend invites to the team members beyond PIs- people actually doing much of the work - who have few opportunities to engage.

  - Chair needs to be able to coordinate the event and have professional event management and conferencing people to execute the event (they can be funded via the conference proposal).
  - Recommend that the event is held in the city of the chair(s)
  - Find way to get PI + CoPIs for outreach/mailing list
  - Setup long-term DN?

- **Enabling the ML/AI Transformation of Science Discovery and Innovation**(see sec: 5.1.3)
  - Investments leading to low barrier access to high end CI
  - Support innovative approaches to training on ML/AI methodologies, workflow development, and infrastructure usage.
  - Investments in standardized data formats and robust frameworks to enhance data quality and interoperability
  - Develop programs that address the need for long-term support for personnel
  - Ensure that AI resources are inclusive and accessible to a broad range of researchers

- **Access to hardware resources**(see sec: 5.2.4 for details)
  - Introduce Small, EAGER-like Awards for Hardware Maintenance and Upgrades
  - Enhance Support for Scalable Data Storage Solutions
  - Expand Training Programs for Heterogeneous Computing and Interoperability
  - Develop Campus-Level "Condo Model" for Shared Hardware Resources
  - Fund Studies on Improving AI Hardware Energy Efficiency
  - Establish a Long-Term Hardware Support Fund
  - Develop a Sustainable Data Management Strategy
  - Establish Permanent NSF-Funded Training Centers
  - Promote Regional or National "Condo Models" for Shared Resources
  - Prioritize Energy Efficiency in Hardware Funding Decisions: Make energy efficiency a key

- **Role of RSEs and Support Staff** (see 5.3.4 for details).
  - Diverse funding models for stable support
  - Invest in RSEs integral to major CI investments
  - Match RSE expertise and experience to domain science
  - Support teams of RSEs with different levels and domains of expertise

- **Designing Specific AI Tools for Science Discovery and Innovation** (see 5.4.4 for details)
  - Early career AI training opportunities
  - Support for Shared data and sharing mechanisms
  - Infrastructure support for domain specific AI tools

- **Training Resources** (see 5.5.2 for details)
  - Train research computer scientists to work with domain experts (mainly grad students). Domain experts know how to compute pre-HPC (e.g. Office applications on a laptop), and need help to use HPC with AI.
  - Train research software engineers to be experts in computer science and familiar with the target domains.
  - Develop a workforce that spans undergraduate students, who are not experts in the domain-science, but, who are nimble at learning software tools and can provide maintenance for installing needed packages
  - Teach community to overcome silos in goal setting (performance vs actual research goals.) and set realistic objectives
  - Train users to identify the limitations of "black-box" AI: what worked, why, etc
  - Delivering of training - look to non-traditional ways that training is deployed
  - Provide a searchable mechanism within domain-specific areas so others can learn more quickly and identify good tools to use
  - Organize frequent workshops targeting PIs, domain scientists, RSEs, and computer scientists to learn novel methods and keep up with evolving AI trends

- **Translating Science Advances into CSSI Tools: from Papers to Software** (see 5.7.4 for details)
  - **Encourage mechanisms for providing credit for software and tools.** Require DOI's for publicly available software/data resources; citations and papers in appropriate journals. Encourage inclusion of tools and their usage in grant reporting across both CSSI and domain science grants to promote recognition of software development activity.
  - **Sustainability grants for software tools for science.** The need grants for sustainability to keep software working was reinforced. Recent addition of the sustainability track was recognized. Such support "keeps the lights on and provides a base for rapid adoption and sustained use of computing driven science.
  - **Emphasize science culture with reproducibility of science advances and reuse of available scientific tools.** Reproducibility initiatives that promote widespread sharing of scientific tools based on science advances help.
  - **Balanced funding sources.** Certain funding sources should prioritize students while others support research scientists and engineers who can develop and maintain professional quality software.
  - Bridging the gap between community stakeholders and tool developers is crucial for interdisciplinary collaboration. There is often a lack of recognition for these efforts and the challenges of balancing specialized tools with broader generalization. Tools may not be immediately appreciated by the community, and reliance on student teams poses risks if they are not sustained beyond graduation. Funding priorities sometimes favor students over research scientists and engineers, impacting the development and maintenance of high-quality software. Long-term support, potentially modeled after National Labs, is important for scaling and updating software. Globus is highlighted as an example of effective software support through a freemium model that balances open access with premium features.
- **Integrating AI/Foundation Models into CSSI** (see 5.8.4 for details)
  - To address the challenges of using foundation models for cyberinfrastructure for sustained scientific innovation, there are two primary recommendations.
  - there needs to be clear domain-specific standards for data collection, data curation and data sharing. The resulting data can be used to train foundation models for use in CI.
  - Second, there needs to be a substantial investment in developing this data infrastructure, and for training individuals to maintain, contribute to, and use this data infrastructure.
- **Integrating New Hardware into CSSI Software** (see 5.9.4 for details)
  - Change the perception of POSE / CSSI-sustainability grants from "end of funding" to "next step in maintaining a healthy software base".
  - Create systems in reporting to value software like papers, track the requested metrics around adoption, etc., and hold PIs *accountable* for them (e.g. impacts on renewal or future SW funding).
  - Spread the OAC culture of valuing software across NSF.
  - Create a clearer path to transition CSSI-created software products throughout the NSF CI ecosystem (ACCESS, LCCF, etc.).
  - Can we work with institutions to create an ecosystem to help support research software? (e.g. underlying expectations in release procedures, network infrastructure, etc.).
- **Community Building & Measurable Broader Impacts** (see 5.10.4 for details)
  - Refine and Expand Metrics: Collaborate with social scientists to develop new metrics that capture both quantitative and qualitative aspects of community engagement. Consider what tools can analyze community sentiment, diversity, and long-term engagement trends. Allow communities to define what matters to them.
  - Innovate Broader Impact Strategies: Encourage original thinking in broader impact planning. Ensure appropriate budget allocations for long-term community engagement and sustainability initiatives. Explore the use of unstructured surveys to gain deeper insights.

- Strengthen Privacy and Legal Compliance: Work with legal experts to ensure all tools and methods used in community-building are compliant with Federal, State and privacy laws. Consider adopting practices like partial name storage or anonymization to protect personally identifiable information while maintaining functionality. Community leads must make a habit of requesting consent for all personal information collected and leveraging their institutional review boards (IRBs).
- Invest in Expertise: There is a need for a coordinated effort that brings expertise together in a cohesive manner. Mechanisms could include hiring or consulting with social scientists or other specialists in community engagement to guide the development of community-building initiatives. Provide training and resources to investigators to help them better understand and implement these strategies.

• **Sustainability & Continuing Training** (see 5.11.6 for details)

- Structured Organization and Community Building: Develop an 'Alliance' model which fosters broader partnerships, bringing projects together under collaborative umbrellas, similar to Research Coordination Networks (RCNs), BigData Hubs, AI institutes. Alliances and hubs, which may be discipline or regionally focused, will share resources and best practices via a central repository to avoid duplication of efforts. Regional collaboration can foster the sharing of cybertraining-proficient people for scalability. Create a digital library of repository materials and share formal curricula through this federated repository.
- Community-driven and Innovative Curriculum: The cyberinfrastructure community must adapt to rapidly changing technology by ensuring that materials and delivery methods remain up to date and relevant. Community-driven curriculum and topics are important to maintain our competitive advantage. Programs similar to the Software Carpentries program should be considered to teach cyberinfrastructure to scientists, specifically cloud based infrastructure as a service. Programs should expand their target audiences to include K-1, community colleges and current CI professional from academia and industry. Programs should be considered which teach professional skills in addition to technical skills, include effective communication, project management and cybersecurity/privacy awareness.
- Inter-directorate and Industry Collaboration: It was suggested that the development of two-way relationships with other NSF directorates, such as EDU and domain directorates, would be useful, specifically incorporating EDU pedagogy research into CyberTraining awards. A strong recommendation from this year's participants and last year's participants is to partner with the NSF TIP directorate and industry partners to develop training which is valuable to industry for staff upskilling providing an opportunity for subsidies and sustainability in addition to other opportunities.

1.4 Broader Impacts

The PI meeting brought together leading experts in the CSSI and broader communities to discuss and share innovations and best practices of developing and sustaining cyberinfrastructure over time. These new and reinforced collaborations will build the capacity for sustainable cyberinfrastructure services that can enhance productivity and accelerate innovation in science and engineering and will significantly contribute to increasing the impact of the output of NSF and specifically OAC's programs. The sharing of best practices around focused panel discussions and posters will allow many impacts outside the primary disciplines for which the awards had been made.

Careful consideration was made in selecting speakers and panelists to ensure they came from a diverse set of backgrounds consistent with community best practices.

2 Meeting Overview, Goals, Planning and Execution

2.1 Meeting Goals

The CSSI PI meeting supports the community building efforts of past Software Infrastructure for Sustained Innovation (SI2) and CSSI workshops. CSSI PI meetings provide a forum for PIs to share technical information about their projects with each other, NSF program directors and others; to explore innovative topics emerging in the software and data infrastructure communities; to discuss and learn about best practices across projects; and to stimulate new ideas of achieving software and data sustainability and ensuring a diverse pipeline of CI researchers and professionals. PIs also provide valuable feedback to the program on emerging opportunities and challenges. For this year in addition to the orginal goals of building and sustaining good cyberinfrastructure we also expected to see the effect of the AI driven transformation on the NSF CSSI/Cybertraining research community. The PI meetings have resulted in the formation of many new collaborations along with a sharing of best practices.

2.2 Committee and Meeting Organization

2.3 Meeting Content

The final program provided a strong overview of CSSI, Cybertraining, SCIPE and had representation from other NSF/OAC programs like ACCESS and CC* (see program in Figure 2.3). A keynote, two sets of shared panels (for all participants) and one set of parallel panels (one for CSSI and one for Cybertraining) were complemented with two sets of breakouts and three sets of poster sessions(required for every funded project).

A keynote talk by the director of NSF/OAC anchored the program and provided a comprehensive overview of NSF/OAC programs and plans [1]. The panels provided active discussion from a distinguished set of invited community leaders (see details in subsequent sections). The breakouts provided an opportunity for every participant to provide input that we have processed into findings and recommendations that are very insightful not only into successes and pain points of existing programs but also provide some guidelines for future interests of participants. The posters (open access in figshare; see appendix for list of posters) are at the heart of this meeting. The level of participation was extremely high and the poster rooms were busy and hosted many productive conversations supporting networking and future collaborations.

3 Community Feedback

We conducted an anonymous post-meeting survey of all attendees to get feedback about different aspects of the meeting, including pre-meeting communication. We received 163 responses from the 284 registrants (some of whom were unable to attend the meeting), which is a response rate of at least 57%. We had 116 responses from those affiliated with the CSSI program; 44 responses from those affiliated with the Cybertraining program; 10 responses from those affiliated with the SCIPE program; and the remaining responses came from individuals affiliated with other programs.[2]

Our respondent's roles at the meetings and roles at their institutions are summarized in Figures 2 and 3.

[1] Travel disruptions moved this talk to a later time to no detriment in the overall flow of the meeting

[2] These values include overlaps as individuals can be affiliated with more than one program.

Program NSF CSSI, Cybertraining and other OAC programs PI Meeting

Monday, August 12, 2024

8:00 – 9:00	**Registration (Preconvene --- Outside of Symphony Ballrooms) and Breakfast (Symphony Ballrooms I, II, III)**			
	Poster 1 Session Set-up Time			
9:00 – 10:00	**Opening Session**			
	Welcome	Carolina Ballroom	15 minutes	A. Patra/ M. Thomas Co-Chairs
	Keynote talk	Carolina Ballroom	30 minutes	K. Antipas, NSF/OAC Direct
	Q&A (s)	Carolina Ballroom	10 minutes	
	Transition to Posters	Carolina Ballroom		
10:00 - 11:00	**Poster Session 1 and coffee**	Symphony (IV), V, VI, VIII		
11:00 – 11:45	**Shared Panel (Panel I) Topic: CI for AI In Science: Role of NAIRR**	Carolina Ballroom		
	Moderator: Dan Stanzione, (TACC/UT Austin)			
	Panelists: Katie Antypas (NSF), Varun Chandola(NSF), Ben Brown (DOE/OASCR), D. K. Panda(OSU)			
11:45-12:30	**Breakouts Session I**			
	I. 1 Enabling the ML/AI based transformation of science discovery and innovation (Abani)	Mecklenburg I		
	I. 2 Access to hardware resources (Alan)	Mecklenburg II		
	1. 3 Role of RSEs and other support staff (Jeff)	Mecklenburg III		
	I. 4 Designing specific AI tools for science discovery and innovation (Vivak)	Governor's Ballroom 1 & 2		
	I. 5 Training resources for PIs (domain scientists) and RSE/computer scientists (Mary)	Governor's Ballroom 3 & 4		
	I. 6 Enabling the ML/AI based transformation of science discovery and innovation (Shaowen)	Governor's Ballroom 5 & 6		
12:30 – 13:45	Lunch	Symphony I, II, III		
14:00 – 14:45	**Parallel Panels (Panel II)**			
	CSSI			
	Moderator: Abani Patra	Mecklenburg I, II, III		
	Topic: Future-Proofing Investments for CSSI Tools	Sustaining tools over changing hardware and software landscapes		
	Panelists: Ian Foster (U. Chicago/ Argonne), Dan Negrut (UW Madison), Daniel Crawford(Virgina Tech), Mike Heroux(Sandia)			
	Cybertraining/SCIPE			
	Topic: Community expe	Carolina Ballroom		
	Moderator: Alan Sussman	Carolina Ballroom		
	Panelists: Manish Parashar (Utah), Dave Hart (NCAR), Ritu Arora (Wayne State), Mary Ann Leung (Sustainable Horizons Institute)			
14:45 - 15:00	Break			
15:00 - 15:45	**Breakouts Session II**			
	CSSI:			
	II.1 Open Science and CSSI (Guoyu)	Mecklenburg I		
	II.2 Translating science advances into CSSI tools (from papers to software) (Abani)	Mecklenburg II		
	II. 3 Integrating AI/Foundation Models into CSSI (Vivak)	Mecklenburg III		
	II.4 New hardware and CSSI software (Dan Stanzione)	Governor's Ballroom 1 & 2		
	Cybertraining:			
	II.5 Community building & broader impacts (should discuss metrics) (Dhruva)	Governor's Ballroom 3 & 4		
	II. 6 Sustainability & Continuing Training (Jane)	Governor's Ballroom 5 & 6		
15:45 - 16:00	Break			
16:00 – 17:00	Report Outs and Open Discussion (Leads will present results from breakouts)	Carolina Ballroom		
5:00 – 7:00	**Posters 2**	Symphony (IV), V, VI, VIII		
	Reception			

Tuesday, August 13, 2024

8:00 – 9:00	Registration and Breakfast		
9:00 – 9:45	**Shared Panel (Panel III)**		
	Moderator: Jeff Carver	Carolina Ballroom	
	Topic: The People Resource Gap		
	Panelists: Dan Katz (UIUC), Sandra Gesing (US-RSE), Michael Zentner (UCSD/SDSC), Chuang Wang (UNCC)		
9:45 – 10:30	**OAC Resources and Program Highlights**		
	CC*: Amy Apon (NSF), Scotty Strachan (Nevada System of Higher Education)	Carolina Ballroom	
	ACCESS: David Hart (NSF NCAR)		
	OAC Core / CDS&E: Sheikh Ghafoor (NSF)		
10:30 – 11:30	**Posters 3 and Coffee**	Symphony (IV), V, VI, VIII	
11:30-12:00	***Parallel Event I***		
	Cybertraining-SCIPE	Carolina Ballroom	
	Topic: TBD		
	Speaker: Sharon Geva (NSF)		
	Parallel Event II		
	(Organizing Commitee and Invitees Only) Work on report from panels & breakout sessions	Executive Boardroom (Second Floor)	
12:00-12:30	PIs are welcome to have collaboration meeetings for their respective projects	Anywhere	
12:30 – 1:30	Lunch	Symphony I, II, III	
13:30 - 15:30	NSF Office Hours		
		Marlon Pierce	Mecklenburg I
		Varun Chandola	Mecklenburg II
		Sharon Geva	Mecklenburg III
		Sheikh Ghafoor	Governor's Ballroom 1 & 2
		Amy Apon	Governor's Ballroom 3 & 4

Figure 1: Program and Schedule

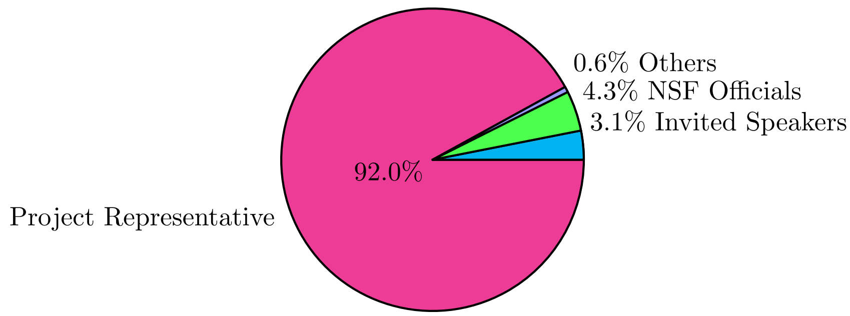


Figure 2: Primary role of survey respondents at the meeting

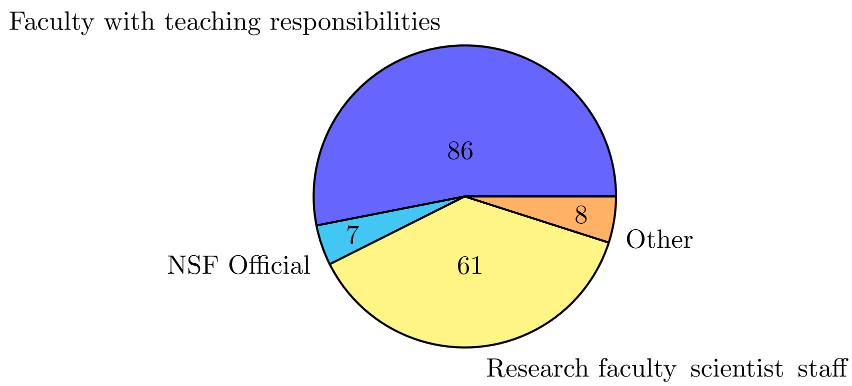


Figure 3: Primary role of survey respondents at their home institution.

3.1 Feedback on Meeting Organization

About 93% of respondents were neutral or found the meeting organization positive. From their written feedback on meeting organization, respondents said the short time frame for announcements resulted in issues with planning travel and accommodation, and had conflicted with other personal commitments. Some respondents experienced additional financial costs because of the short timeframe. Otherwise, respondents were generally pleased with the communication about the meeting.

3.2 Feedback on Usefulness of the Meeting

Respondents were asked to provide their disposition towards different components of the PI meeting as either positive, neutral or negative. The results are presented in Table 1. Respondents were also allowed to provide textual feedback.

From the text feedback, the poster sessions were the most positive experience of the PI meeting with some respondents suggesting that there be more poster sessions and that they be held for longer periods of time to allow for more interaction and discussion. Many respondents highlighted the social events and networking opportunities as critical parts of their experience as it gave them opportunities to meet colleagues across disciplines for future collaboration. The breakout sessions received mixed feedback: respondents wanted more time and more structure for the breakout sessions. However, as we see in section 5 many constructive ideas were contributed in the breakouts for the future development of the program.

Event Category	Positive	Neutral	Negative
Poster Sessions	147	16	1
Social Events	134	28	2
Networking / Informal Engagements	128	35	1
NSF Program Highlights	113	43	8
Panels	89	61	14
Breakout Sessions	71	64	29
NSF Office Hours	65	91	8

Table 1: Disposition of respondents towards different components of the PI meeting.

3.3 Feedback on Meeting Program and Events

Most attendees were pleased with the format of the meeting and did not express any specific changes. The remaining attendees' feedback is summarized as follows.

1. Attendees wanted more time for interaction, including opportunities for structured networking, sessions to interact with other PIs on best practices and potential for collaboration, and more time during poster sessions.
2. Attendees wanted more time for NSF office hours as only a few slots were available per program officer in attendance.
3. Attendees were not as engaged with panel sessions and preferred more targeted discussions in smaller groups.
4. Attendees found the breakout sessions too broad in focus and too short in time. They wanted clearer goals for the sessions.
5. Attendees found the focus on AI/ML too strong in panels and breakouts, leaving them feeling left out if their topic did not coincide with AI/ML.
6. Some attendees suggested an event specific to first time NSF PI Meeting attendees.

3.4 Feedback on Meeting Location and Timing

Most attendees were pleased with the location with the meeting in Charlotte, NC as it was a central hub for air travel. However, they found the venue to be isolated from nearby restaurants and cafes, making it difficult for them to socialize after the meeting period. Many attendees suggested Washington D.C., Chicago and Denver as desirable meeting locations as these places are easily accessible or centrally located in the country.

The timing of the meeting for arriving on a Sunday and departing on a Tuesday were generally positively viewed. Respondents suggested that August 10 to August 12 of next year would be an ideal time to have the meeting.

3.5 Feedback on the Costs of Attending the Meeting

Respondents were split on the registration fee: some found the fee reasonable, while others expressed dissatisfaction with a registration fee for a mandatory meeting that is funded by the NSF. Respondents also commented on the high hotel rate.

3.6 Summary Recommendations Based on Feedback

- **Location:** should be central and easily accessible (e.g. Washington D.C., Chicago and Denver) with affordable hotel rates.
- **Timing:** Mid-August Sunday-Tuesday timing was satisfactory though earlier notice will be helpful for planning travel.
- **Additional time for discussion and networking:** PIs expressed need for additional time for networking and structured/unstructured time to learn best practices.
- **Panels and Breakouts:** Panels attracted poor response while the breakouts were felt to be useful but rushed and not fine grained enough. Recommend reducing panels and using the time saved for more focused breakouts.
- **NSF PD Office Hrs:** Additional office hours if possible are recommended.

4 Panels and Keynotes

During the meeting, there was one Keynote Speaker session (on Day 1) and four Panel sessions, spread across Days 1 and 2. Details follow in the sections below.

4.1 Keynote Speaker Summary

The Keynote address, titled "Update on NSF's Office of Advanced Cyberinfrastructure and the National AI Research Resource Pilot," was presented by Katie Antypas, Director, Office of Advanced Cyberinfrastructure. A copy of her presentation can be found here: [PDF]

The keynote provided the PIs with a great overview of NSF/OAC current and future investments and thinking for the future. Keynote was well attended and inspired/ informed much discussion among attendees. Such a keynote that is effectively a "State of the Programs and Division" address is very helpful and serves well to set the tone of the meeting.

4.2 Panel Session Summaries

The panels were designed to address topics relevant to both the CSSI and the CyberTraining/SCIPE communities. They were organized around a topic, and panelists were asked to answer a set of questions. The panel sessions were then followed by the Breakout sessions (Section 5), where the community met in several small groups and worked together to answer questions related to the panel topics.

Two of the panel sessions were jointly attended by both the CSSI and CyberTraining/SCIPE participants, and one session was dedicated to the separate NSF programs. The Panel sessions are listed in Table 4.2.

4.2.1 Panel I (Joint): CI for AI In Science: Role of NAIRR

Moderator: Dan Stanzione, (TACC/UT Austin)
Panelists: Katie Antypas (NSF), Varun Chandola(NSF), Ben Brown (DOE/OASCR), D. K. Panda(OSU)
This panel explored challenges and opportunities in provisioning CI for the AI revolution in scientific discovery and engineering innovation. AI for science requires CI comprised of hardware, software, data and people resources acting in concert. Core questions addressed included:

Table 2: Panel session details

Panel	Topic	Moderator	Panelists
Panel I: Joint	CI for AI In Science: Role of NAIRR	Dan Stanzione, (TACC/UT Austin)	Katie Antypas (NSF), Varun Chandola(NSF), Ben Brown (DOE/OASCR), D. K. Panda(OSU)
Panel IIa: CSSI	Future-Proofing Investments for CSSI Tools: Sustaining tools over changing hardware and software landscapes	Abani Patra, (Tufts U.)	Ian Foster (U. Chicago/ Argonne), Dan Negrut (UW Madison), Daniel Crawford(Virgina Tech), Mike Heroux(Sandia)
Panel IIb: CyTr/SCIPE	Community experiences and Evolving Needs	Alan Sussman (U. Maryland)	Manish Parashar (Utah), David Hart (NCAR), Ritu Arora (Wayne State), Mary Ann Leung (Sustainable Horizons Institute)
Panel III: Joint	The People Resource Gap	Jeff Carver (U. Alabama)	Dan Katz (UIUC), Sandra Gesing (US-RSE), Michael Zentner (UCSD/SDSC), Chuang Wang (UNCC)

1. What are the major gaps in current CI available to support AI driven science? In particular, what aspects of existing CI designed for more traditional compute intensive sciences need to be expanded.
2. What are challenges in training and preparing the workforce needed?
3. What are the current resources and opportunities for investigators?

4.2.2 Panel IIa (CSSI): Future-Proofing Investments for CSSI Tools

Moderator: Abani Patra (Tufts University)
Panelists: Dan Katz (UIUC), Sandra Gesing (US-RSE), Michael Zentner (UCSD/SDSC), Chuang Wang (UNCC)
Questions for Discussion:

- Sustaining tools over changing hardware and software landscapes
- Collaboration between domain scientists and computer/ software engineers
- HPC Resources are shifting towards GPUs, TensorCores, etc. How do we cope?

4.2.3 Panel IIb (CyTr/SCIPE): Community experiences and Evolving Needs

Moderator: Alan Sussman
Panelists: Manish Parashar (Utah), David Hart (NCAR), Ritu Arora (Wayne State), Mary Ann Leung (Sustainable Horizons Institute)
Questions for Discussion:

- Metrics for Cybertraining/SCIPE programs and outreach
- Large Institutes - how do they get created?
- Community Building/Broader Impacts:
  - Create broader communities, reach out to MSIs; grow new PIs
- Promoting Better Scientific Software

- New Ways/How to multiply the NSF investment – delivering diverse courses, certificates, programs, different domains
- How to effectively broaden the science/engineering research workforce, beyond a one size fits all approach

4.2.4 Panel III (Joint): The People Resource Gap

Moderator: Jeff Carver
Panelists: Dan Katz (UIUC), Sandra Gesing (US-RSE), Michael Zentner (UCSD/SDSC), Chuang Wang (UNCC)
Topic: The People Resource Gap - RSEs, Data/AI scientists
Discussion Topics:

- Topic 1 - Pipeline
  - Driving Question: In your opinion, how should we design training programs to draw people into RSE and related careers? Including:
    - Training on specific skills (including software engineering and other skills)
    - Training programs that integrate multiple skills into a coherent program
    - Structure of training programs (e.g. short courses, apprenticeships)
    - Format of training programs (e.g. online, in-person, asynchronous video)
    - Credentials
- Topic 2 - Professional Development
  - Driving Question: How can organizations best support the career development of people in RSE and other related roles? Including:
    - What types of professional development do they need?
    - Are these roles careers in themselves or a step along the way to something else?
    - How can people in these roles advance within their organizations?
    - What are the impediments to advancement?

5 Breakout Session Summaries

Breakout sessions following keynotes provided a key forum for seeking input from the attending PIs and project representatives. Each breakout was moderated and key questions were posed to start discussions. Notes taken at the sessions formed the basis for the session summaries below. Where applicable, summaries include the following sections: Overview; Current Status/ Challenges; Desired Outcomes and Future Directions; Recommendations.

Original notes can be found in the meeting google drive Breakout Session folder: https://drive.google.com/drive/u/0/folders/1SUKTkqF5I3_gHl7jR_pM8KMpdO64cELH

5.1 Enabling the ML/AI Transformation of Science Discovery and Innovation

Embedding machine learning (ML) and artificial intelligence (AI) methods into scientific tools presents a range of challenges and opportunities. Breakout participants have provided thoughtful input to define primary challenges, desired outcomes and recommendations to enable this transformation.

5.1.1 Current Status/ Challenges

- At the heart of these challenges is the diversity and complexity of scientific data. Scientific datasets are often heterogeneous, coming from various sources and in different formats, which complicates their integration and processing for ML/AI models.
- Additionally, high-quality, labeled data, which is crucial for effective ML/AI training, is frequently incomplete or noisy in scientific fields.
- Another significant challenge is scalability. Fields like genomics and climate modeling generate massive datasets that require substantial computational resources and sophisticated algorithms to manage effectively.
- This issue is compounded by a lack of interdisciplinary expertise; many domain scientists are not well-versed in ML/AI, which limits their ability to leverage these technologies fully.
- Moreover, the transferability of ML models across different disciplines is problematic, as best practices and models can vary widely, making it difficult to select the most appropriate approach for specific scientific problems.
- The validation and interpretability of ML/AI results are also critical concerns. For ML/AI to be useful in scientific research, results must be validated and provide interpretable insights into fundamental science questions.
- Additionally, infrastructure and standardization issues further complicate the integration of ML/AI into scientific workflows. The difficulty in switching between different AI tools, such as PyTorch and TensorFlow, hampers collaboration and adoption, while the lack of standardized frameworks and processes affects consistency and usability.
- Preparing data for ML/AI applications adds another layer of complexity. The preprocessing of data can be labor-intensive and time-consuming, with issues related to data quality and metadata complicating the process further.

5.1.2 Desired Outcomes / Future Directions

A primary desired outcome is a skilled and well trained workforce capable of enabling and sustaining the ML/AI transformation. Therefore, investing in enhanced training and education for domain scientists is crucial. In the future we there will be available training on ML/AI methodologies for both specialists and domain scientists and accessible workflows and infrastructure to significantly improve our ability to utilize these technologies. Additionally, support for integrating ML-based courses into graduate programs can help build foundational knowledge and skills.

Low barrier access to high end CI (computing, data, workflows and training harnesses) that allows rapid development and exploration will maximize science advances and innovation. Thus, improving infrastructure and resources is a key priority. Leveraging initiatives like the National AI Resource Research (NAIRR) can provide access to necessary computing resources and promote the standardization of AI tools and models. Developing AI consulting services similar to HPC consulting could also help researchers integrate ML/AI best practices into their work.

Solutions in place for data management and interoperability for effective ML/AI integration. Promoting standardized data formats and frameworks can enhance data quality and facilitate interoperability between different AI tools. Improving data pipelines for preprocessing and integration will make data more accessible and useful for ML/AI models. Investing in hybrid solutions that combine AI with traditional computational methods can help tackle domain-specific challenges more effectively.

Fostering community and collaboration is vital for advancing ML/AI in scientific research. Encouraging cross-disciplinary collaboration between computer scientists and domain experts can lead to the development of domain-specific AI solutions and methodologies. Creating benchmarking metrics and evaluation datasets tailored to scientific domains will also help in validating and comparing AI models.

Sustainability and long-term investments are crucial for maintaining progress in AI research. Addressing the need for long-term funding for personnel and developing career paths for engineers and developers are important steps.

Lastly, ensuring that AI resources are inclusive and accessible to a diverse range of researchers can enhance the utility of AI approaches. By addressing these challenges and leveraging the identified opportunities, CSSI and related communities can better integrate ML/AI methods into tools for scientific discovery and innovation.

5.1.3 Recommendations

The following recommendations summarize the outcomes of the discussion:

- **Investments leading to low barrier access to high end CI** (computing, data, benchmarks, workflows and training harnesses) to enable rapid development and exploration will maximize science advances and innovation.
- **Support innovative approaches to training on ML/AI methodologies, workflow development, and infrastructure usage.** Such training in multiple modalities – workshops to classes for both domain scientists and specialized CI personnel like research software engineers on a priority basis at scale where adequate numbers of personnel at every major research institution has access to such training.
- **Investments in standardized data formats and robust frameworks** to enhance data quality and facilitate interoperability between different AI tools.
- **Develop programs that address the need for long-term support for personnel** and career paths for engineers and developers who enable the development of AI/ML resources.
- **Ensure that AI resources are inclusive and accessible to a diverse range of researchers** from different geographic regions and demographics.

5.2 Access to Hardware Resources

5.2.1 Background

We opened with a discussion of what we mean by hardware when it comes to CI. Hardware includes traditional CI resources (e.g. high-performance computing); however, In the modern AI era, the notion of hardware is augmented with a heterogeneous mix of specialized accelerators (FPGAs, GPGPUs [of course], chiplet designs, and other novel systems found only in specialized labs (e.g. Cerebras, Graphcore, SambaNova, others). Beyond computation, storage remains a challenge, where hosting a multiple TB or PB scale dataset remains a challenge, including reliable long-term backup and permanent hosting/dissemination of large datasets to enable reproducible science. Beyond hardware, there are the ongoing costs to keep the systems running (maintenance, system/network admins) and to help users (research engineers/RSEs).

5.2.2 Current Status/Challenges

The following represent key pain points/challenges when it comes to hardware to support CI research:

- Data Storage and Migration: The inefficiencies and time waste associated with migrating data across different locations are significant impediments.

- Co-Located Compute and Storage: The lack of co-located compute resources with storage creates bottlenecks in processing efficiency.
- Software Compatibility and Interoperability: Heterogeneous computing introduces software incompatibility issues, necessitating better support for interoperability frameworks.
- Training and Personnel: There is a clear need for more training resources and skilled personnel to support these systems.
- Awareness and Access to Resources: Researchers are often unaware of available CI resources, and existing portals do not make access sufficiently user-friendly.
- Scalable Storage Solutions: AI's storage needs are vast, requiring more scalable solutions that include long-term availability and artifact preservation.
- Energy Efficiency and Long-Term Costs: As hardware obsolescence slows, energy efficiency and the long-term costs of running AI hardware become critical considerations when determining whether a resource is obsolete and is worth the ongoing power and maintenance costs.
- Flexible and Sustained Funding Models: There is a need for more flexible and sustained funding models that support both short-term maintenance and long-term hardware investments.
- Cloud and On-Premises Balance: While cloud computing offers flexibility, it can be costly for experimentation, highlighting the need for a balanced approach between cloud and on-premises resources. Nevertheless, any on-premises or national CI resources should offer a compelling option to what is possible in the cloud as many researchers are opting for commercial solutions.

5.2.3 Desired Outcomes / Future Directions

If the recommended changes were implemented, several scientific use cases would flourish, particularly in fields requiring large-scale data processing and long-term storage. Environmental modeling, genomic research, and AI-driven simulations in physics are prime examples where access to flexible, scalable storage solutions and high-performance computing resources is critical. These projects often generate massive datasets that need to be processed efficiently and stored securely over extended periods, something that academic infrastructure can provide more effectively than commercial clouds. The flexibility of academic environments allows for customized hardware setups, such as heterogeneous computing platforms and the "condo model" for shared resources, which can be optimized for the specific and evolving needs of these complex research endeavors.

Beyond just cost and efficiency, this flexibility fosters creativity in how researchers approach their work. Sensitive research areas like genomics or social science studies benefit from the ability to implement stringent data security measures and develop long-term data management strategies within dedicated academic data centers, rather than relying on the more rigid frameworks of commercial clouds. Moreover, the freedom to tailor computing environments to meet the unique demands of interdisciplinary AI research or unconventional data processing enables researchers to explore new methodologies and push the boundaries of their fields. By investing in these adaptable and creative infrastructure solutions, NSF-funded projects can avoid the limitations of commercial cloud dependency, such as unpredictable costs and challenges with long-term data accessibility, while enhancing the innovative potential of scientific research.

5.2.4 Recommendations

We propose the following short-term recommendations:

- **Introduce Small, EAGER-like Awards for Hardware Maintenance and Upgrades:** Create a mechanism within NSF for small, targeted awards that allow for the maintenance and incremental upgrades of existing hardware throughout a project's lifecycle (even beyond the grant period). This would help ensure that equipment remains functional and up-to-date, minimizing disruptions between awards and performance issues.

- **Enhance Support for Scalable Data Storage Solutions:** Provide supplemental funding specifically aimed at high-capacity data storage solutions. This could involve integrating on-premises storage with cloud-based systems to better meet the large-scale demands of AI-driven research projects.
- **Expand Training Programs for Heterogeneous Computing and Interoperability:** Increase funding for workshops and training programs focused on heterogeneous computing environments and software interoperability. These programs should target both students and researchers, equipping them with the skills needed to effectively manage and utilize diverse hardware and software systems.
- **Develop Campus-Level "Condo Model" for Shared Hardware Resources:** Offer grants that support the adoption of a "condo model" for shared hardware resources on campuses. In this model, centralized resources are managed at the institutional level, while individual research groups have the option to add dedicated nodes, optimizing resource use and reducing costs.
- **Fund Studies on Improving AI Hardware Energy Efficiency:** Provide funding for empirical studies and pilot projects that focus on retrofitting existing AI hardware to improve energy efficiency. This would help reduce the long-term operational costs associated with running AI-driven research, making it more sustainable.

We propose the following long-term recommendations:

- **Establish a Long-Term Hardware Support Fund:** Develop a dedicated fund within NSF grants that supports the full lifecycle of hardware, from acquisition to decommissioning. This should include provisions for ongoing maintenance and energy efficiency improvements, ensuring long-term usability and sustainability of research infrastructure.
- **Develop a Sustainable Data Management Strategy:** Implement a long-term strategy for data storage and management that includes the preservation of research artifacts and datasets. This could involve the creation of NSF-funded data centers dedicated to ensuring that data generated by NSF-funded projects remains accessible and usable over time. Leverage what other agencies know about managing extreme data, e.g. data storage from high-energy physics (DOE), and ingesting all internet traffic (NSA).
- **Establish Permanent NSF-Funded Training Centers:** Create NSF-funded training centers that offer ongoing support and education in heterogeneous computing and software engineering/interoperability. These centers could also serve as hubs for the development and testing of new frameworks, ensuring that researchers are equipped with the latest tools and knowledge.
- **Promote Regional or National "Condo Models" for Shared Resources:** Encourage the development of regional or national "condo models" for hardware resources, where multiple institutions share access to high-performance computing resources. This would reduce costs per researcher and increase access to state-of-the-art technology across the academic community.
- **Prioritize Energy Efficiency in Hardware Funding Decisions:** Make energy efficiency a key criterion in NSF's hardware funding decisions. By prioritizing equipment and infrastructure that offer better energy performance, the NSF can help ensure that AI research remains sustainable both financially and environmentally.

5.3 Role of RSEs and Other Support Staff

5.3.1 Background

This topic focuses on the roles that various software-facing CI Professionals (i.e. Research Software Engineers, Data Scientists, and Research Infrastructure Engineers) play in the research ecosystem, how those roles may evolve in the future and what support the community needs to facilitate these changes.

5.3.2 Current Status

Across domains and across institutions, it is clear that there are a wide range of roles that RSEs play in STEM research & education, both in terms of the kinds of work that they do and the formal positions they hold. Examples included:

- Faculty/PIs and graduate students who need computational tools or novel software for their scientific work, self-teach or train for the skills needed, and create purpose-built software on their own.
- Research labs where a lab member becomes the primary specialist in software and coding skills, serving as a resource either formally (a software engineer or other technical expertise, hired for software support) or informally (a graduate student or postdoc who specializes and teaches others).
- Full-time RSEs who work full-time, funded as staff on a single grant, often housed in an academic department, but sometimes a university's central RSE institute. These staff may have PhDs (or otherwise have deep domain specialist knowledge), coupled with more informal/on-the-job software skills. Or, they may have a more traditional software background and then learn to work with scientific collaborators.
- Software infrastructure engineers that are operationally focused via software development in support of research, education, and instruction. These are more like modern IT professionals in that they have application services that they maintain, defined in software, that are related to networking, security, identity, and cloudy infrastructure.
- Attendees from NCAR & UCSD represented national centers with full-time RSEs who work on rotating portfolios of projects with scientists.

Across contexts, many attendees spoke about the challenges of finding and retaining staff for these roles, which require project-specific combinations of software engineering, scientific, and communication/project management skills. People in these roles are often faced with communication barriers between researchers and engineers, who are largely trained to describe projects in quite different terms. Engineers are most effective if they are given clear requirements but researchers don't want to commit the time to developing them (and suffer negative career consequences if they do). Software engineering teams are seen as current "high value/high impact" for research-facing activities in academia, but are also the hardest to retain.

5.3.3 Desired Outcomes / Future Directions

Discussion groups spoke extensively about the need to provide formal support for RSE time and expertise, and to develop career models that make sense for various contexts, whether that be dedicated career tracks focused on computational work and software engineering, recognizing faculty and research staff efforts, or working with university-level centers such as libraries and computing centers. These roles need to be elevated from "support" to coequal to research staff, closely integrated and integral to making the science happen - through both career recognition, citations, as well as stable career paths, in order to recruit/retain these professionals away from industry. Some important features might include ensuring recognition of software projects as first-class scholarly outputs, and some level of autonomy for RSEs to pursue their own funded research.

Another thread of discussion focused on the need to spend the time to understand what kinds of software expertise and projects are needed where, and what kinds of resources provide the most benefit to different communities of researchers. Differentiating between software applications/workflows and software-as-infrastructure is also important. One attendee summarized: "The needs for specialized services become more clear/focused at departmental scale, this is where RSE's and other CI professionals are the most impactful for individual science domains." Others mentioned that many emerging research institutions or departments may need dedicated attention to fundamental data engineering and data management capacities before getting to more complex or high-compute software workflows.

5.3.4 Recommendations

Diverse funding models for stable support Attendees recognized a need for a variety of different funding models, combining federal grants, institutional funding, and potential fee-for-service and cost-sharing models (e.g. NSF funds first 3 years, university funds after that). For sustainability, granting agencies might need to consider both larger grants for building out significant new capabilities, as well as determining what ongoing needs are required - just as with other physical infrastructure, software infrastructure entails maintenance and ongoing support for users. The shape and scope of collaborations also needs to be able to vary - in many cases, critical software infrastructure is maintained by a small group of people at one or two institutions, but used by a community of scientists from many different institutions.

Invest in RSEs integral to major CI investments Both universities and funding agencies are encouraged to invest in RSE staff as a key component in the research enterprise, just as they do with computing or with large hardware investments like MRI, telescope facilities, etc.

Match RSE expertise and experience to domain science Discussion also focused on maintaining appropriate recognition for the diversity of contexts and needs that software-facing CI professionals operate across. The field needs support for the work it takes to find the right matches between RSE and projects where DE is useful, and for researchers to take the leap of learning to communicate with software engineers. At the department and domain level, the kinds of interactions that are most important may vary - the resources and projects needed probably will not look the same in chemistry as in political science.

Support teams of RSEs with different levels and domains of expertise Different domains and different institutions will also involve a continuum of maturity or professionalization of RSE work. In some cases, research institutions need foundational Data Engineering professionals (management, transport, staging, curation), who may be housed in libraries, software carpentry type training resources, or dedicated software centers, etc. - before specialized software can be effectively used. In others, teams that combine RSE skills with domain expertise can be embedded in departments or research institutes to serve portfolios of projects that benefit their fields.

5.4 Designing Specific AI Tools for Science Discovery and Innovation

5.4.1 Background

Artificial Intelligence (AI) has come to the forefront as an area of national research focus. AI holds promise to aid and even transform science discovery and innovation, yet AI is still in its infancy. How AI tools can be used, which ones should be developed, and how they should be deployed is an evolving preoccupation of the research community.

5.4.2 Current Status

Current AI use cases suggest insight into these questions. For instance, AI is being used to predict dangerous algal blooms in freshwater nearly two weeks prior to their appearance based on heterogeneous data from solar radiation measurements, water oxygenation levels, precipitation records and nutrient loads in freshwater. AI is also being used to improve parameter exploration in high-fidelity fluid dynamics simulations for a variety of atmospheric and geophysics applications.

5.4.3 Desired Outcomes / Future Directions

While such applications of AI underscore its potential utility, there are many areas of science discovery and innovation where AI tools have yet to make a substantial impact.

Potential Use Case 1. Binary exploitation uses a computer's standard operations, often memory control behaviors, to gain access to the computer's data or processes. Often, binary exploitation comes with delivering a payload to the target system to encourage certain types of behavior. The development of heterogeneous payloads such as text and images is an area that is ripe for AI tools.

Potential Use Case 2. Scientific workflows often require working with a variety of structured and unstructured materials ranging from published documents, metadata and hand-written lab notebooks. As research challenges become more complex, these scientific workflows also become increasingly complex and often distributed amongst multiple scientists and researchers. To support efficient, rigorous collaboration, scientific workflows can benefit from AI tools that can amalgamate, search through, relate and provide access to a variety of relevant research materials.

5.4.4 Recommendations

To realize these potential opportunities for AI tools in science discovery and innovation, a number of coordinated efforts and investments are needed.

Early career AI training opportunities From a workforce development perspective, students need early training in AI/ML tools, including their development and utilization; and individuals need to be encouraged to pursue specialized careers in AI/ML, computer science and domain fields with stable financial support.

Support for Shared data and sharing mechanisms From a scientific rigor and democratization perspective, researchers need to be incentivized to develop more shared data and shared cyber-infrastructure resources; researchers need mechanisms to facilitate agreements with private groups and government agencies around data use and harmonization; and researchers need support and training for working with private or high-security data sets.

Infrastructure support for domain specific AI tools From a coordination perspective, the community needs federal-level guidance on priorities; support for archiving, accessing and using data and infrastructure after project lifecycles; and investments into cross-disciplinary efforts to develop AI tools for specific domain problems.

5.5 Panel Topic Breakout Session: Training Resources for PIs (Domain Scientists) and RSE/Computer Scientists

Description This panel explored challenges and opportunities in provisioning CI for the AI revolution in scientific discovery and engineering innovation. AI for science requires CI comprised of hardware, software, data and people resources acting in concert.

Objective The goal of this breakout is to explore the need for (if it exists), and the challenges and opportunities for developing AI/NAIRR relevant training resources for PIs (domain scientists), RSEs, and computer scientists.

5.5.1 Discussion Summary

In this session, we were asked to answer three questions:

How do we define the NAIRR Workforce? Including PIs (domain scientists), RSEs, and computer scientists, more? The group came up with the following list of job categories or titles: Educators, Engineers, Facilitators, Faculty, Master, Post-docs, Professors, Staff, Undergrads, Users Academic, Domain-science, Experts, Multi-disciplinary, Pipeline, Policy, Teaching.

We then created a list of descriptions for these jobs:

- Research scientist (faculty, scientist, post-docs, grad students) for target domains.
- Research software engineers as facilitators of technology.
- Teaching faculty, educators, and librarians for training and educational resources.
- One person may have different hats.
- Domain experts that need computer or data scientist support and knowledge.
- Domain experts that need to use/learn AI.
- Research computer scientists that work with domain experts (mainly grad students). Domain experts know how to compute pre-HPC (e.g. Office applications on a laptop), and need help to use HPC with AI. Research computer scientists need with the transition.
- Need research software engineers that are experts in computer science and familiar with the target domains.
- It is hard to keep trained personnel without competitive salaries as compared to industry.
- Still need high-performance computing experts.
- Teaching faculty have the potential to be trained on how to train the research scientist and research software engineers.
- Traditional RCD positions
- Data scientists together with domain scientists
- Responsible AI, ethics, data wrangler, and data governance
- Legal consultant - acquisition from entities, security
- Policy makers
- Evaluations (for different parts of the pipeline / workflows)
- We focussed on the academic aspect of the workforce, as that is what we are familiar with.
- The workforce spans undergraduate students, who are not experts in the domain-science, but, who are nimble at learning software tools and can provide maintenance for installing needed packages
- more advanced undergrads/master students who can research best existing software tools and guide undergrads
- Postdocs (also staff and professors) who are multi-disciplinary experts in CS / ML-AI / domain science and need to understand
- researchers, perhaps using AI tools.
- domain scientists, computer scientists, software engineers users of resources such as GPUs/FPGAs
- trainers, co-ordinators
- pipeline of students/postdocs
- policy makers defining the use of data

What are the challenges for developing AI/NAIRR relevant training resources for PIs (domain scientists), RSEs, and computer scientists?

- Overcoming silos in goal setting (performance vs actual research goals.), remove barriers for interdisciplinary goals.
- Scalability with goals, students don't have time to optimize code, better to run something now than optimize later. Need RSEs knowledge, importance of communication between RSEs and domain experts.
- Hard for domain scientists to learn the limitations of the AI "black-box" tools, and bring previous experience to new workflows.
- Hard to know why a AI model worked, "explain" why the model produced the result that it did. Amplified with LLM, how can it be trusted?
- Interdisciplinary training. Bridging between the domain science the data science / AI
- Defining the training needs based on a new science
- Consolidate tools - easy to use / easy to access
- Providing relevant data
- Which model/algorithm to apply to a specific application - better filter mechanism of tools
- Mismatch in context: what level of fake rate is acceptable.
- Excited students might see AI as a silver bullet, and may not understand training data (eg. molecular dynamics simulations) may be very expensive.
- As a domain expert, the most difficult issue is talking a different language from computer science. Experience solving it with a bootcamp between RSEs and domain scientists to find a common language.
- Really steep learning curve. Harder to train because it is new.
- Hard to switch tools once a tool is
- Need knowledge of the target domain.
- Access to the scalable AI Infrastructure
- Proprietary hardware - Portability
- Language barriers - Data Scientist use one language / domain uses another language
- Pre-train or foundation models
- Lack of policies for defining where and when to use AI
- Rapid turn-over in technologies (software & hardware).
- Students are excited about AI, but don't know which kinds of models are appropriate, how to apply them, data needs, etc.
- Language/semantic barrier. Need someone to translate, including understanding which tool might be useable vs one that can not meet target precision needs
- AI is multidisciplinary; teaching AI reaches beyond the students' domain more.
- we need to target different audiences/groups; not one size fits all
- most of the data, resources, and algorithms are currently proprietary.
- Reliance on commercial models for training materials
- technology is moving too quickly. Things are becoming obsolete in a matter of months.
- Update existing training material/courses to be current.
- Should PhD students be focussed on domain or AI tools?
- Retention issues of well trained students/researchers to industry.
- Domains such a social science/public policy/medicine have not usually used AI techniques, computer techniques, so there is a large gap.

What are the opportunities for developing AI/NAIRR relevant training resources for PIs (domain scientists), RSEs, and computer scientists?

- Bootcamp to develop common languages, interdisciplinary teams. Bootcamp a good framework to bring people from different domains.
- Different domains may want to collaborate in creating the training resources. The same problems are solved over and over again, and this only gets discovered when talking between disciplines.
- Make resources searchable, ask for support and collaborate with librarians that know how to organize,and make resources discoverable.
- NSF TIP Directorate framework, partner with industry to create knowledge hubs.

- Different workshops targeting PIs, domain scientists, RSEs, and computer scientists to get novel work and get the AI trends
- Many of the software tools can be taught generically, and not needed by local PIs/postdocs
- partnership with industry. Students will be used to using such tools. Industry would like to lock-in students.
- Open source community at the university level.
- We could add domain-specific list of tools useful for a given domain, explanation tools, eg. that might recommend something useful for visualization
- Webinars that can showcase solutions found, short-videos for a particular skill. The challenge is to break it down in digestible chunks, but still discoverable and relevant.
- Similar efforts for High Performance Computing have been accomplished.
- Delivering of training - look to non-traditional ways that training is deployed
- Partnering with people that use AI
- Inclusive training - having a robust catalog of training available for the community
- Younger levels in education (AI) - high schoolers, to provide teaching opportunities for them - fundamental AI
- Collaborative meetings between domain scientists, computer scientists, RSEs to develop unified AI tools.
- Funding agencies seem to be aware for the need of domain-specific RSE funding
- We can provide a curated list - that is well known and well communicated to the community - of good AND bad, online learning tools for various tools
- We could add domain-specific list of tools useful for a given domain, explanation tools, eg. that might recommend something useful for visualization
- If Open Science requirements are made more stringent by funding agencies, eg. required publishing of software, installation instructions etc., it would provide a searchable mechanism within domain-specific areas how others can learn more quickly good tools to use
- They could be mandated to do this? But this could have its risks.

5.5.2 Recommendations

- Train research computer scientists to work with domain experts (mainly grad students). Domain experts know how to compute pre-HPC (e.g. Office applications on a laptop), and need help to use HPC with AI.
- Train research software engineers to be experts in computer science and familiar with the target domains.
- Develop a workforce that spans undergraduate students, who are not experts in the domain-science, but, who are nimble at learning software tools and can provide maintenance for installing needed packages
- Teach community to overcome silos in goal setting (performance vs actual research goals.) and set realistic objectives
- Train users to identify the limitations of "black-box" AI: what worked, why, etc
- Delivering of training - look to non-traditional ways that training is deployed
- Provide a searchable mechanism within domain-specific areas so others can learn more quickly and identify good tools to use
- Organize frequent workshops targeting PIs, domain scientists, RSEs, and computer scientists to learn novel methods and keep up with evolving AI trends

5.6 Community building & measurable broader impacts

Objective The objective of this breakout was to discuss ideas for developing communities in a sustainable manner and measure the impacts of their work.

5.6.1 Discussion Summary

Questions The discussion was seeded with the following questions:

- Can metrics be embedded in how we build communities?
- How do you typically measure broader impacts in your community?
- What would you like to do differently in how broader impacts are considered?
- Are communities better off being integrated with curricular materials?
- How should one contextualize metrics for broader impact?
- What are the privacy issues one should consider?

Background Broader impacts extend from advancing scientific research and societal outcomes. Broader impacts achieved by building communities is a core tenet of the NSF CSSI, Cybertraining and SCIPE programs. These three programs target different outcomes, and as such build and support communities with different focuses and interests. While this offers several possibilities in advancing Broader Impacts and societal outcomes, meaningfully measuring their outcomes continues to remain a challenge for these programs.

Current Status The National Science Foundation (NSF) expects researchers' work to have broader impacts: the potential to benefit society and contribute to the achievement of specific, desired societal outcomes. Practices promoting broader impacts have come a long way. Perhaps we are approaching a set minimum expectations.

The NSF does not want to be prescriptive about the societal outcomes a project addresses but provides examples of broader impacts across several categories. These include inclusion, STEM education, public engagement, societal well-being, STEM workforce, partnerships, national security, economic competitiveness, infrastructure. As such, the communities in question can take different forms. Among others, the broader impacts in these communities can be achieved via developer communities, research support structures, curricular adoption, and informal training. Practices to enhance the broader impacts in a community are typical today. For example, transitioning ad hoc (or informal) training to the classroom and sharing curricular materials are increasingly considered standard practices. Today, they need to be considered with the same intentionality as science. It is not sufficient to merely mention activities. When considering broader impacts, researchers should consider the objectives, the activities, and the budgetary considerations.

Researchers are actively collecting data on extended research outcomes such as publications, software developed, students graduating, funding success, and courses developed. We note that while there are public tools, they may be hard to use because they might expose student names, FERPA issues.

Depending on their scope and funding, communities can take years to build.It is hard to gauge the success of community development over the lifetime of a shorter grant. Collecting metrics is somewhat easier in software development projects. Smaller programs are not appropriately funded to perform a true evaluation. Some schools have teams that can assist with evaluations and collect metrics.

5.6.2 Desired Outcomes / Future Directions

The programs in this cohort are developing communities that focus on three critical albeit different aspects of advancing computationally enabled research. The CSSI program is developing communities of developers, and researchers who can adopt them in their workflows. The Cybertraining communities focus on the development and adoption of training materials in informal and formal settings. The SCIPE programs is focused on developing communities that can support researchers effectively use computing technologies.

The communities in this cohort will have some overlapping needs, support structures, and ambitions. With this in mind, these communities can morph into others. They will further impact other communities downstream. As such, metrics have to consider the impact on secondary and tertiary communities. The community notes that these impacts may come into play after the duration of the original program. A prescriptive mechanism that tracks the outcomes and objectives of a community could delineate its progres, but runs into the danger of becoming restrictive.

Reporting on the Broader Impacts is ingrained in most communities. While these three programs try to capture metrics such as demographics, attendance, publications from their communities, there is a need for more enhanced metrics. Participants in these communities engage for different reasons. While some communities are transient, others may exist for longer. One may consider how we can measure the sense of belonging in a community. Perhaps a mechanism that supports a longitudinal study could capture this information. There is an opportunity to elevate the discussion on what we are collecting. For example, there is a deeper need for mechanisms that help researchers develop, adopt and share new scientific practices. Unlike research outcomes, it is harder to capture this need in metrics. While we are focused on advancing the use of metrics in a community, mechanisms to protect the participating researchers' privacy, and intellectual property need to be considered as well. It is possible that other programs have developed unique ways to address this challenge. Here, specialized expertise for creating and evaluating approaches might prove to be helpful. A group structured to meaningfully assist researchers engaging in these communities collectively could be helpful.

5.6.3 Recommendations

Refine and Expand Metrics Collaborate with social scientists to develop new metrics that capture both quantitative and qualitative aspects of community engagement. Consider what tools can analyze community sentiment, diversity, and long-term engagement trends. Allow communities to define what matters to them.

Innovate Broader Impact Strategies Encourage original thinking in broader impact planning. Ensure appropriate budget allocations for long-term community engagement and sustainability initiatives. Explore the use of unstructured surveys to gain deeper insights.

Strengthen Privacy and Legal Compliance Work with legal experts to ensure all tools and methods used in community-building are compliant with Federal, State and privacy laws. Consider adopting practices like partial name storage or anonymization to protect personally identifiable information while maintaining functionality. Community leads must make a habit of requesting consent for all personal information collected and leveraging their institutional review boards (IRBs). Invest in Expertise There is a need for a coordinated effort that brings expertise together in a cohesive manner. Mechanisms could include hiring or consulting with social scientists or other specialists in com munity engagement to guide the development of community-building initiatives. Provide training and resources to investigators to help them better understand and implement these strategies.

5.7 Translating Science Advances into CSSI Tools: from Papers to Software

5.7.1 Background

The CSSI program has traditionally enabled the computational science community to translate new scientific innovations in modeling and related ideas into tools for widespread community usage. [3] Our group here presented their thoughts on the principal challenges associated with this goal of CSSI. The session focused on addressing the question "What are the main challenges to developing and sustaining new community tools for enabling and representing new science advances?"

5.7.2 Current Status

- Getting credit and recognition for developing community software and tools continues to be a challenge. Public resources for sharing software like github are increasingly useful and reduces the problem for students going to industry but academic recognition is still an ongoing challenge. Introduction of new journals like Journal of Open Scientific Software (JOSS) [4] provide mechanisms for getting metrics like citations to support traditional measures of academic success. However, the core problem of recognition remains since many software tools for science are intrinsically harder to generate and such measures of credit limited. The core workforce of computational and domain scientists need recognition to create a sense of ownership and accomplishment for success. Tool and software development is often perceived as a career killer for science personnel as a consequence of this lack of recognition.
- Maintenance and sustainability of new tools is really hard – getting money for updating and maintaining software is really hard. In most domain getting resources for hiring and supporting Computing Science personnel remains a challenge even in support of big instruments (e.g. LIGO). Support for this staff is critical but not understood and rarely budgeted for though elements of programs like the sustainability track of CSSI are helping. More permanent positions that are not dependent on single tools are needed for engineers who need to build careers.
- Furthermore, just keeping software tools functioning is inadequate since new scientific advances need to be continuously assimilated to keep tools scientifically relevant. Developing new tools for minor advances and/or similar models that can benefit from the same programming and data infrastructure is wasteful but often the "novelty" of the science often makes it easier to find support for such development. Summarizing, a principal challenge to developing tools that translate new computational science into tools that make it seamlessly available to the community remains sustained funding.
- A persistent challenge is the need for effectively delivering tools to domain scientists. Well organized communications and community building and outreach through workshops, town halls and such are needed. Platforms like github now provide mechanisms for sharing but community needs to be built and sustained. Progress on community specific platforms and science gateways(see for e.g. [5, 6]) with extensive investments through the CSSI programs and domain science specific programs has been very good. Many such platforms are doing well though few have solved the sustainability challenges. [6] designated as a software institute has created much support and programming for science gateways serving large range of science communities.
- Software tools especially those that have complex usage modalities may take a long time for uptake. We need to bring community stakeholders and tool developers together - bridging the domain/CI gap. We need to evaluate the future value of the work and sustain it through the uptake cycle. Tools that meet a pressing need for the community tend to be more popular. However, it is not always possible to make a tool flexible enough to meet the needs of the community. Thus, generalizing a niche tool used for an advance is important. The tension between specialization (useful in specific context) vs. generalization (usefulness to broader community) is one that each community and tool need to develop.
- While we started this section highlighting the insufficient acknowledgement / recognition of tool development efforts, insufficient recognition of the interdisciplinary efforts needed here are also often lost.
- Much scientific software still has a dependence on the student teams that may not last beyond students' graduation dates for creating tools based on new innovations (often in the student's dissertations). Such tools are often inadequately developed for scaling to more users.

5.7.3 Desired Outcomes/ Future Directions

Software developers and other enablers for computing and data driven science should be first class citizens of the research ecosystem.

The MOLSSI (one of two software institutes that were funded) focuses on computational chemistry tools and ha built a healthy ecosystem of domain scientists, software engineers and support personnel and trains UG and Grad students to learn to write sustainable code and launch into community. This is an excellent model.

The desired stable state of processes for effective translation of science advances to tools requires both infrastructure(hardware, tools and people) and culture. A principal part of the culture is a focus on reproducibility in science done with computing.

Sustainability of tools with a variety of resources to support the needed people and platforms is a much desired goal. Globus is an example of a widely used, open access software, that is supported via a model that works: freemium with premium features available to subscribing institutions.

Successful CSSI tools often have long-term champions (e.g. the Department of Energy National Labs support a number of tools – e.g. Trilinos and PetSc) who bring in funding, personnel and the continued intellectual inputs needed.

5.7.4 Recommendations

- **Encourage mechanisms for providing credit for software and tools.** Require DOI's for publicly available software/data resources; citations and papers in appropriate journals. Encourage inclusion of tools and their usage in grant reporting across both CSSI and domain science grants to promote recognition of software development activity.
- **Sustainability grants for software tools for science.** The need grants for sustainability to keep software working was reinforced. Recent addition of the sustainability track was recognized. Such support "keeps the lights on and provides a base for rapid adoption and sustained use of computing driven science.
- **Emphasize science culture with reproducibility of science advances and reuse of available scientific tools.** Reproducibility initiatives that promote widespread sharing of scientific tools based on science advances help.
- **Balanced funding sources.** Certain funding sources should prioritize students while others support research scientists and engineers who can develop and maintain professional quality software.

Bridging the gap between community stakeholders and tool developers is crucial for interdisciplinary collaboration. There is often a lack of recognition for these efforts and the challenges of balancing specialized tools with broader generalization. Tools may not be immediately appreciated by the community, and reliance on student teams poses risks if they are not sustained beyond graduation. Funding priorities sometimes favor students over research scientists and engineers, impacting the development and maintenance of high-quality software. Long-term support, potentially modeled after National Labs, is important for scaling and updating software. Globus is highlighted as an example of effective software support through a freemium model that balances open access with premium features.

5.8 Integrating AI/Foundation Models into CSSI

5.8.1 Background

Foundation models have become extremely popular with the public owing to their humorous natural language capabilities and with developers because of their development into programming assistance tools. Owing to these features, determining whether foundation models can benefit cyberinfrastructure for sustained scientific innovation has become an important concern.

5.8.2 Current Status

Several existing applications of foundation models for cyberinfrastructure suggest that there is a potential benefit of such models to CI for scientific innovation.

For example, foundation models for materials science and chemistry learn a representation of the underlying chemistry so that a point in this representation captures relevant physical chemical data of the system. This representation can be used to build other cyberinfrastructure for downstream applications in materials science and chemistry, including simulations and experimental design.

Foundation models are also used to represent complex, large-scale genomic information. In turn, these foundation models can be used to monitor and predict behaviors of concerning pathogens, such as predicting SARS-CoV-2 variants prior to their emergence.

5.8.3 Desired Outcomes / Future Directions

These successes suggest that foundation models can be part of, spur new, or help build, novel cyberinfrastructure for sustained scientific development. Examples of potential use cases and future directions are listed below, the desired benefit of realizing these use cases, and the challenges to achieving them.

Potential Use Case 1. As hardware continues to evolve and computing systems become more heterogeneous, software often needs to be modified and updated to ensure that they can take advantage of these capabilities to deliver faster or more energy efficient results. However, writing such software is rather challenging and requires developers to not only know how to exploit the hardware, but also requires them to know the scientific domain to ensure the integrity of results.

Owing to the difficulty of this task, foundation models that can write high-concurrency code for heterogeneous architectures automatically can be very valuable. Such a foundation model would dramatically improve the performance of software and ensure its sustainability even as hardware continues to change.

Challenges to achieving this use case include developing adequate data for training such models; ensuring that the results are correct; and ensuring that hardware capabilities are being exploited.

Potential Use Case 2. Cyberinfrastructure for science depends on a clear understanding of the literature in the scientific domain. As manuscript and publication rates grow, CI development needs support for analyzing developments in the literature.

Developing foundation models for this task can help developers stay on top of advancing research and identify opportunities for improving CI.

Challenges to achieving this use case include how this data should be collected, processed and referenced by the foundation models, especially in light of copyrighted materials that are mediated by publishers.

5.8.4 Recommendations

To address the challenges of using foundation models for cyberinfrastructure for sustained scientific innovation, there are two primary recommendations. First, there needs to be clear domain-specific standards for data collection, data curation and data sharing. The resulting data can be used to train foundation models for use in CI. Second, there needs to be a substantial investment in developing this data infrastructure, and for training individuals to maintain, contribute to, and use this data infrastructure.

5.9 Integrating New Hardware into CSSI Software

Objective
How can we more effectively bring CSSI-produced software to the wide group of users the ACCESS/HPC Resource Providers reach? What are the barriers to integration?

Suggested Questions
See report template below, which can help guide question development.

- Why isn't your CSSI software product available through ACCESS?
- Should it be? Is this an outcome you want?
- What can NSF do? What can the Resource Providers do?
- How big a challenge is supporting constantly changing hardware platforms for you as a software provider?

5.9.1 Background

A persistent challenge across CSSI tool sustainability is adapting the constant innovation in computing hardware. In recent years for example this has encompassed the transition to the universal use of GPUs for providing the bulk of the computing resources in HPC. New hardware clearly adds costs to supporting software as do updates in the broader software environment, vendor releases, and related factors. Strategies for sustaining and maintaining CSSI-supported software is the core issue addressed in this breakout.

5.9.2 Current Status

Summary themes from breakout:

- There is no clear path to get SW out through the RPs - find people you know at each one and ask is the status quo.
- Can NSF help by raising accountability and reporting?
  - Like PAR for SW
  - OAC values SW... other divisions don't! Get to that "first class" Ben Brown proposed – grants are all "new", so track record in successful SW doesn't matter.
  - If we want good SW, hold people accountable for those usability metrics.
- Link more clearly to papers for discoverability
- POSE/Sustainability track
  - In a grant to develop SW, 3 years long, you barely have a viable product before the end, and yet you are expected to push it, do outreach etc.
  - Those grant tracks are perceived as "terminal funding"
- Churn of HW/support is painful (postdoc lost 6 months with supercomputer transition at NCAR)
  - Though OS upgrades can be as painful as new HW.
- Even if we solve all this, networking, data, Inf. Sec. costs still will be huge problems.
  - One site could not release their software open source without clearing expensive institution-mandated information security reviews.
- Where should the institution boundary be on solving these problems? Do we make it worse that every project crosses institutional boundaries?
- How do we balance demands on institutions with the need for diversity of institutions (i.e. avoid a system of "haves" and "have nots").

5.9.3 Desired Outcomes / Future Directions

Create incentives for people to write better software, and reward the creation and maintenance of good software throughout government.

5.9.4 Recommendations

- Change the perception of POSE / CSSI-sustainability grants from "end of funding" to "next step in maintaining a healthy software base".
- Create systems in reporting to value software like papers, track the requested metrics around adoption, etc., and hold PIs *accountable* for them (e.g. impacts on renewal or future SW funding).
- Spread the OAC culture of valuing software across NSF.
- Create a clearer path to transition CSSI-created software products throughout the NSF CI ecosystem (ACCESS, LCCF, etc.).
- Can we work with institutions to create an ecosystem to help support research software? (e.g. underlying expectations in release procedures, network infrastructure, etc.).

5.10 Community Building & Measurable Broader Impacts

Objective This breakout was focused on community building,and measures of the resulting broader impact with a careful description of current states, future impacts and recommendations to attain desired goals. [3]

5.10.1 Background

Broader impacts extend from the advancement of scientific research and societal outcomes. Broader impacts achieved by building communities are a core tenet of the NSF CSSI, Cybertraining, and SCIPE programs. These three programs target different outcomes and, as such, build and support communities with different interests and goals. Although this offers several possibilities in extending Broader Impacts and societal outcomes, meaningfully measuring their outcomes continues to remain a challenge for these programs.

5.10.2 Current Status

The National Science Foundation (NSF) expects researchers' work to have broader impacts: the potential to benefit society and contribute to the achievement of specific, desired societal outcomes. Practices promoting broader impacts have come a long way. Perhaps we are approaching a set of minimum expectations.

The NSF is intentionally not prescriptive about the societal outcomes a project addresses, but provides examples of broader impacts across several categories. These include inclusion, STEM education, public participation, societal well-being, STEM workforce, partnerships, national security, economic competitiveness, infrastructure. As such, the communities in question can take different forms. Among others, the broader impacts in these communities can be achieved through developer communities, research support structures, curricular adoption, and informal training.

Practices to improve the broader impacts in a community are typical today. For example, transitioning ad hoc (or informal) training to the classroom and sharing curricular materials are increasingly considered standard practices. Today, they need to be considered with the same intentionality as science. It is not

[3] Notes can be found here: [meeting notes]

sufficient to simply mention activities. When considering broader impacts, researchers should consider the objectives, activities, and budgetary considerations.

Researchers are actively collecting data on extended research outcomes such as publications, software developed, students graduating, funding success, and courses developed. We note that while there are public tools, they may be hard to use because they might expose student names, FERPA issues.

Depending on their scope and funding, communities can take years to build. It difficult to gauge the success of community development over the lifetime of a shorter grant. Collecting metrics is somewhat easier in software development projects. Smaller programs are not appropriately funded to perform a true evaluation. Some schools have teams that assist with evaluations and collect metrics. These facilities are, however, not available to all.

5.10.3 Future Directions

The programs in this cohort are developing communities that focus on three critical albeit different aspects of advancing computationally enabled research. CSSI programs are actively developing communities of developers, and researchers who can adopt them in their workflows. The Cybertraining communities focus on the development and adoption of training materials in informal and formal settings. The SCIPE programs are focused on developing communities that can support researchers effectively use computing technologies.

The communities in this cohort will have some overlapping needs, support structures, and ambitions. With this in mind, these communities can morph into others. They will further impact other communities downstream. As such, metrics have to consider the impact on secondary and tertiary communities. The community notes that these impacts may come into play after the duration of the original program. A prescriptive mechanism that tracks the outcomes and objectives of a community could delineate its progress, but runs into the danger of becoming restrictive.

Reporting on the Broader Impacts is ingrained in most communities. While these three programs try to capture metrics such as demographics, attendance, publications from their communities, there is a need for more enhanced metrics. Participants in these communities engage for different reasons. While some communities are transient, others may exist for longer. One may consider how we can measure the sense of belonging in a community. Perhaps a mechanism that supports a longitudinal study could capture this information.

There is an opportunity to elevate the discussion on what we are collecting. For example, there is a deeper need for mechanisms that help researchers develop, adopt and share new scientific practices. Unlike research outcomes, it is harder to capture this need in metrics. While we are focused on advancing the use of metrics in a community, mechanisms to protect the participating researchers' privacy, and intellectual property need to be considered as well. Other programs may have developed unique ways to address this challenge. Here, specialized expertise for creating and evaluating approaches might prove helpful. A group structured to meaningfully assist researchers engaging in these communities collectively could be helpful.

5.10.4 Recommendations

Refine and Expand Metrics Collaborate with social scientists to develop new metrics that capture both quantitative and qualitative aspects of community engagement. Consider what tools can analyze community sentiment, diversity, and long-term engagement trends. Allow communities to define what matters to them.

Innovate Broader Impact Strategies Encourage original thinking in broader impact planning. Ensure appropriate budget allocations for long-term community engagement and sustainability initiatives. Explore the use of unstructured surveys to gain deeper insights.

Strengthen Privacy and Legal Compliance Work with legal experts to ensure all tools and methods used in community-building are compliant with Federal, State and privacy laws. Consider adopting practices like partial name storage or anonymization to protect personally identifiable information while maintaining functionality. Community leads must make a habit of requesting consent for all personal information collected and leveraging their institutional review boards (IRBs).

Invest in Expertise There is a need for a coordinated effort that brings expertise together in a cohesive manner. Mechanisms could include hiring or consulting with social scientists or other specialists in community engagement to guide the development of community-building initiatives. Provide training and resources to investigators to help them better understand and implement these strategies.

5.11 Sustainability and Continuing Training

5.11.1 Background

The NSF Cybertraining program has been very successful. The program has evolved from supporting training programs for users of CI systems to the introduction of the SCIPE program which focuses on the development of and training for CI professionals. With that success comes the question of sustainability and scalability of the training programs and initiatives which have been supported. The 2023 Cybertraining PI workshop report included a recommendation to further define and recommend strategies for sustainability and scalability of program initiatives which prompted the development of this breakout session.

5.11.2 Objective

Define and recommend strategies for sustainability and the focus of the Cybertraining program for the future.

5.11.3 Current Status

Sustainability and continuation of the NSF Cybertraining and SCIPE program is an important topic and concern of the community. The program has been focused on curriciulum, training (short-term) and NSF directorate priorities. There is an opportunity to expand from small scale projects to a national scope or domain-wide integration and regional collaboration. There are limited industry partnerships and integration with the NSF TIP directorate which is seen as a missed opportunity for sustainability. The majority of the focus has been on college students (graduate).

5.11.4 Suggested Questions

Where has Cybertraining been focused? Where should it focus in the future? Communities (i.e. college students, PIs, CI Professionals), topics, artifacts, curriculum What is the future of Cybertraining? How do we scale, sustain and coordinate, gain efficiencies? What is the role of industry (NAIRR, NSF TIP directorate)? What components are needed for a successful center, institute?

5.11.5 Desired Outcomes / Future Directions

A sustainable resource and personnel ecosystem for training and education of the cyberscientists, RSEs and associated resources without *ad hoc* and one-time grant funding. Training pathways, career definitions and business models for both are well defined and invested in.

5.11.6 Recommendations

Structured Organization and Community Building Develop an 'Alliance' model which fosters broader partnerships, bringing projects together under collaborative umbrellas, similar to Research Coordination Networks (RCNs), BigData Hubs, AI institutes. Alliances and hubs, which may be discipline or regionally focused, will share resources and best practices via a central repository to avoid duplication of efforts. Regional collaboration can foster the sharing of cybertraining-proficient people for scalability. Create a digital library of repository materials and share formal curricula through this federated repository.

Community-driven and Innovative Curriculum The cyberinfrastructure community must adapt to rapidly changing technology by ensuring that materials and delivery methods remain up to date and relevant. Community-driven curriculum and topics are important to maintain our competitive advantage. Programs similar to the Software Carpentries program should be considered to teach cyberinfrastructure to scientists, specifically cloud based infrastructure as a service. Programs should expand their target audiences to include K-1, community colleges and current CI professional from academia and industry. Programs should be considered which teach professional skills in addition to technical skills, include effective communication, project management and cybersecurity/privacy awareness.

Inter-directorate and Industry Collaboration It was suggested that the development of two-way relationships with other NSF directorates, such as EDU and domain directorates, would be useful, specifically incorporating EDU pedagogy research into Cybertraining awards. A strong recommendation from this year's participants and last year's participants is to partner with the NSF TIP directorate and industry partners to develop training which is valuable to industry for staff upskilling providing an opportunity for subsidies and sustainability in addition to other opportunities.

6 Acknowledgements

This workshop was made possible with NSF Awards 435580 and 2434556. We are grateful to NSF for helping to arrange the collaborative CSSI PI and CyberTraining/SCIPE PI meetings. We also thank the following NSF Program Officers who were present: Sonam Ahluwalia, Amy Apon, Sharmistha Bagchi-Sen, Linkan Bian, Varun Chandola, Sharon Geva, Sheikh Ghafoor, Tom Gulbransen, Marlon Pierce, Plato Smith, Ashok Srinivasan, Wen-wen Tung, Rediet Woldeselassie, Sam Xin, May Yuan. We also want to acknowledge the efforts of the following individuals, whose contributions often went above and beyond the call of duty, and this meeting would simply not have happened, including: Susan Rathbun (SDSC); Jordan Wilkinson and Meghan Rodriguez (Tufts); and Hantao Cui (NCSU) for developing and hosting the event website.

References

[1] NSF. CyberTraining Principal Investigator (PI) Meeting. https://confmeet.github.io/ccs2024/, 2024. Accessed: October 1, 2024.

[2] NSF. CyberTraining Principal Investigator (PI) Meeting. https://cybertrainingpi.wixsite.com/meeting2023, 2023. Accessed: November 3, 2023.

[3] NSF. Cyberinfrastructure for Sustained Scientific Innovation (CSSI) Program. https://www.nsf.gov/funding/opportunities/cssi-cyberinfrastructure-sustained-scientific-innovation, 2024.

[4] The journal of open source software. https://joss.theoj.org.

[5] nanohub: Making data and simulation pervasive. https://nanohub.org. accessed July 2025,.

[6] SGX3. A center of excellence and a software institute. https://sciencegateways.org. accessed July 2025.

A Selected Slides & Presentations

A.1 NSF Funding Opportunities

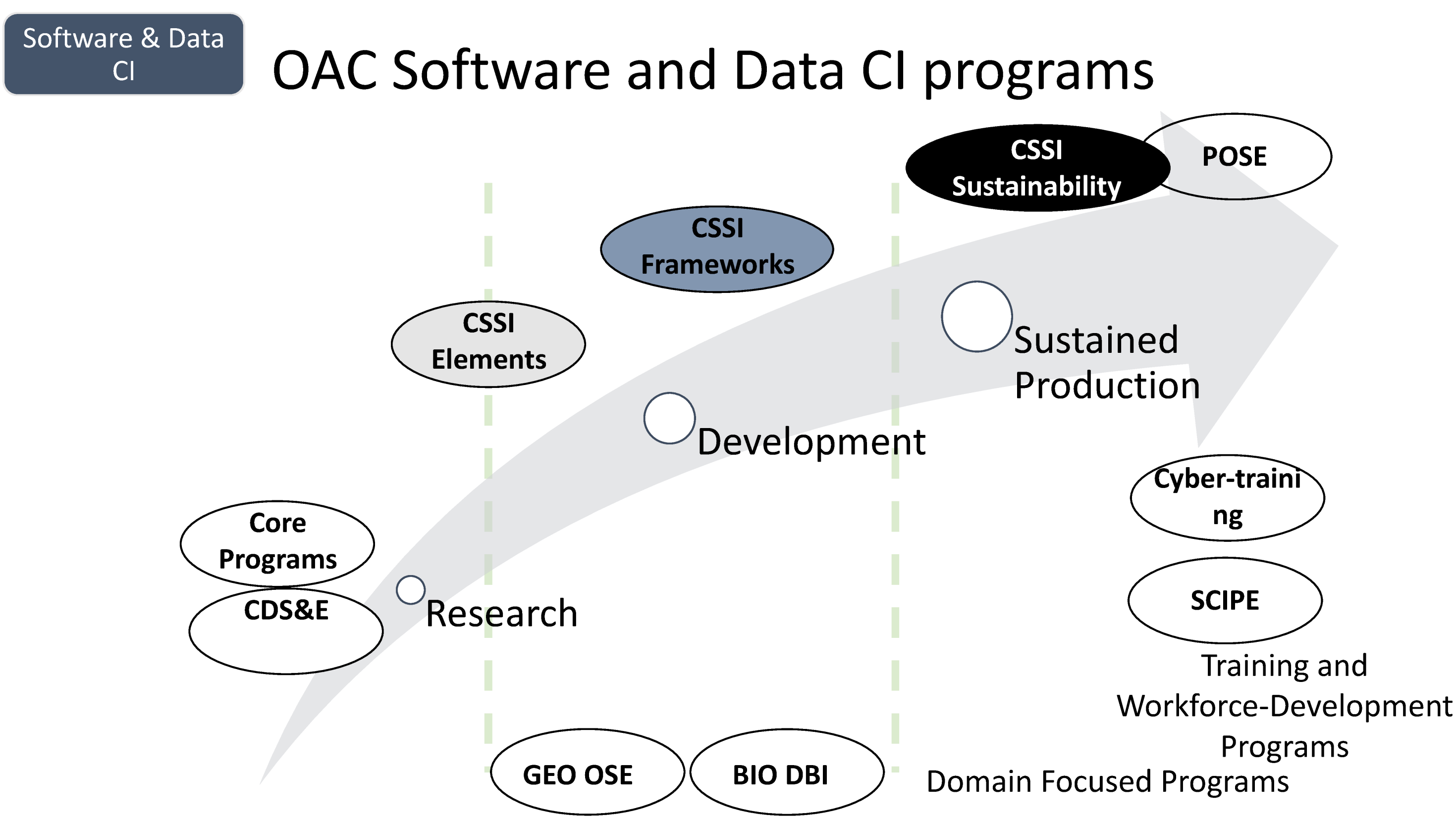
OAC Core, CDS&E, CSSI, and CyberTraining Funding Opportunity

Sheikh Ghafoor, Sharmistha Bagchi-Sen
Office of Advanced CyberInfrastructure

Outline

- Overview of CDS&E, OAC Core and CSSI
- Overview of CyberTraining

OAC Software and Data CI programs



OAC Core

- Innovative proposals for translational research
  - design, development, deployment, experimentation, and application of advanced research CI.
- One or more of the following key attributes:
  - Multi-disciplinary
  - Extreme-scale
  - Driven by science and engineering research
  - An end-to-end solution
  - Deployable as robust research CI
- Small Proposal Only
- Up to ~600k
- **NSF 24-589 Submission Window : October 1 – October 23**

OAC Core: Example Research Topics

- Research in architecture for extreme-scale systems may include design, benchmarking, and analysis of extreme-scale systems for performance, programmability, and usability; storage, networks, and input/output (I/O); data centers and extreme-scale networked systems; and next-generation architectures;
- Research in middleware may include resource management, monitoring, fault tolerance, and cybersecurity;
- Research in scalable algorithms and applications shall be driven by science and engineering applications and may include numerical and high-performance scientific computing methods; data, software, and visualization approaches; and modeling and simulation capabilities; and
- Research in the advanced CI ecosystem may include research in programming languages, libraries, and related environments; performance tuning and interoperability tools; shared CI, e.g., platforms and gateways; and sociotechnical aspects relevant to the advanced CI ecosystem, e.g., best practices, standards, policies, and virtual organizations.

Computational and Data-Enable Science and Engineering (CDS&E)

- ❑ NSF Wide Meta-Program
- ❑ Goal
  - Enabling major scientific and engineering breakthroughs with new computational and data-analysis approaches and best practices.

Developing Essential Methods
- Mathematical, Computational Theories
- Modeling & Simulations

Creating New Paradigms
- Data generation & utilization

Integrating Algorithms & Tools
- Data curation & Analysis
- Visualization

Interfacing between Systems
- Efficient data access & exchanges

Computational and Data-Enable Science and Engineering (CDS&E)

- Encourages research that pushes the envelope of science and engineering
  - Through computation and data,
  - Proposals can be any research area supported by the participating divisions.
- A proposal may address topics that develop or enable interactions among theory, computing, experiment, and observation to achieve progress on intractable science and engineering problems.
- Areas of emphasis for CDS&E vary by program.
- PIs are advised to consult the "related programs“
- **Must have a significant component of computational or data science that goes well beyond what would typically be included in these programs.**
- CDS&E proposal should include substantive science, engineering, or computing research.

CDS&E Participating Divisions

- Directorate for Engineering (ENG)
- Division of Chemical, Bioengineering, Environmental and Transport Systems (ENG/CBET)
- Division of Civil, Mechanical and Manufacturing Innovation (ENG/CMMI)
- Directorate for Mathematical and Physical Sciences (MPS)
- Division of Physics (MPS/PHY)
- Division of Astronomical Sciences (MPS/AST)
- Division of Mathematical Sciences (MPS/DMS)
- Division of Materials Research (MPS/DMR)
- Division of Chemistry (MPS/CHE)

CyberTraining Solicitation Goals

- **Long-term vision:** Computational and Data-driven Science for All scientists and engineers
  - Prepare, nurture, and grow the **scientific research workforce**, including students, instructors, and research CI professionals
- Ensure broad adoption of **CI** tools, methods, and resources
- Integrate CI and CDS&E skills into undergraduate and graduate curricula
  - Address emerging needs and unaddressed bottlenecks through innovative and scalable training
  - Catalyze research with training and educational activities
- **Broaden CI access** and adoption by varied institutions, scientific communities, and underrepresented groups.

Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

- Supports the development and deployment of robust, reliable and sustainable data and software cyberinfrastructure
- Three CSSI Award Classes:
  - Elements: Small groups that will create & deploy robust capabilities for one or more significant areas of S&E (up to $600k, up to 3 years)
  - Frameworks: Larger, interdisciplinary teams around development and application of common infrastructure aimed at common research problems in on or more areas of S&E. (Awards between $600k - $5M, between 3-5 years)
  - Transition to Sustainability: Groups who will execute well-defined sustainability plan that enables new avenues of support for long-term sustained impact of the CI (up to $1M, up to 2 years)
- NSF 22-632 Proposal Deadline: December 1

CyberTraining

Motivation

- **Advanced CI** has a transformative impact on a variety of scientific **research** domains
- The research workforce will benefit from innovative discipline-appropriate training and curriculum materials
- There is a need to foster **broad adoption** of CI resources, tools, and methods by diverse research communities

Project Classes

- **Pilot:** Exploratory projects, $300K over 2 years
- **Small implementation:** $500K over 4 years
- **Medium implementation:** $1M over 4 years

ACCESS coordination
- Share training material in ACCESS Knowledge Base (https://support.access-ci.org/knowledge-base)
- Register expertise in https://support.access-ci.org/cssn

1. Identify challenges in research workforce development
2. (a) Broaden use of CI resources and/or (b) CI skills training – expected to coordinate with ACCESS (access-ci.org)
3. Scalability and sustainability of the training program
4. Recruitment and evaluation plans
5. Collective impact strategy
6. Fostering a suitable community

Pilot

Small

Medium

A.2 NSF ACCESS Resources and Opportunities

NSF ACCESS Resources and Program Highlights

David Hart, NSF NCAR

ACCESS Allocations, co-PI

access-ci.org

Cyberinfrastructure Reimagined

TeraGrid™

2001 – 2011

→

XSEDE

Extreme Science and Engineering Discovery Environment

2011 – 2022

→

ACCESS

Advancing Innovation

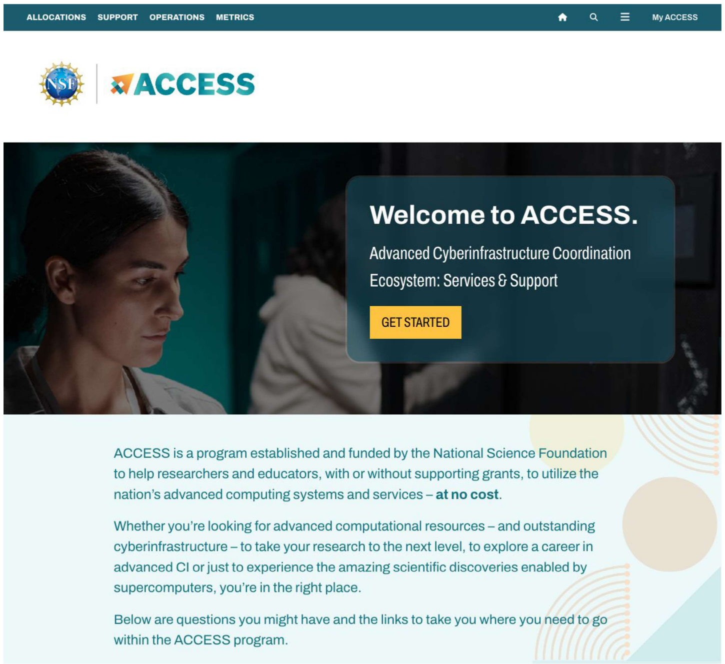
https://access-ci.org

2022 –

NSF ACCESS

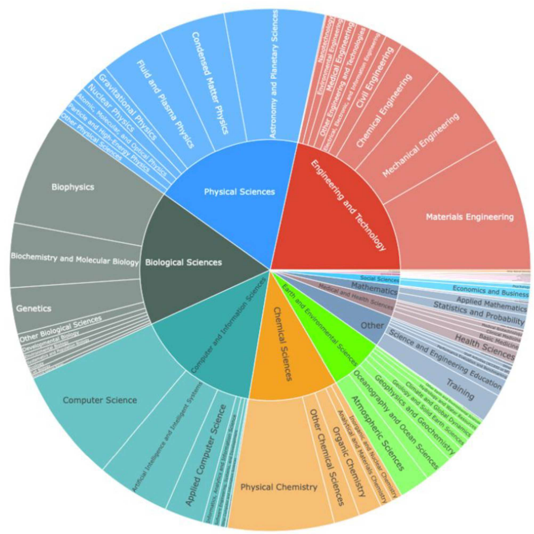
ACCESS has been established by the National Science Foundation (NSF) to help you—the nation's researchers, educators, and CIPs—to benefit from many of the nation's most advanced computing systems and services—all at no cost to you.

Whether it's to take your research to the next level, to pursue a career in advanced CI, or just to explore the scientific discoveries enabled by supercomputers – you have ACCESS.



CI opportunities for

- **Researchers**
  - access to resources to pursue your scholarly objectives in any field of study
- **Graduate students**
  - access to resources and support for your dissertation or thesis
- **Educators / instructors**
  - access for you and your students for courses or training events
- **Resource providers**
  - access to operational services and an extensive community of researchers and educators
- **CI professionals**
  - access to resources, affinity groups, community building, and travel support



Fields of study leveraging ACCESS

Cyberinfrastructure Available

- 30+ Computing systems
  - Varying processors & memory sizes
  - Cloud resources (persistent services)
  - GPUs, vector processors, FPGAs
- Data storage systems
  - Archive, object, tiered
- Software & workflow managers
- High-performance networking
- CI professionals & support tools
- System performance monitoring

Georgia Institute of Technology
Hive Cluster

Indiana University (IU)
Jetstream2

Johns Hopkins University (JHU)
Rockfish

National Center for Supercomputing Applications (NCSA)
Delta

Open Science Grid (OSG)

Open Storage Network (OSN)

Pittsburgh Supercomputing Center (PSC)
Anton 2
Bridges-2
Neocortex

Purdue
Anvil

San Diego Supercomputer Center (SDSC)
Expanse
Voyager

Science Gateways Community Institute (SGCI)
SGX3

Stony Brook University
Ookami

Texas Advanced Computing Center (TACC)
Ranch
Stampede3

Texas A&M University (TAMU)
ACES
FASTER

University of Delaware
DARWIN

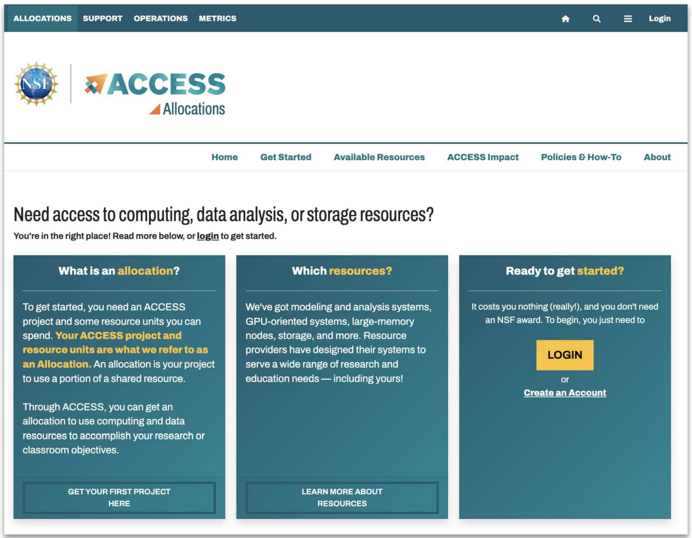
University of Kentucky
KyRIC

Current ACCESS Resource Providers

Open OnDemand available on many ACCESS machines

Delivering an Open, Inviting, and Democratized Resource Marketplace

- Building a welcoming gateway that inspires collaboration and participation
- Providing equitable access across disciplines, institutions, and demographic groups
- Including DEI & continuous improvement in every aspect of our project



Accelerating your time to ACCESS

The average project now takes just over 10 days to go from requesting a project to recording its first use of an ACCESS resource.

Accounts on resources are available in about three days.

Ecosystem Access Time (days)	***2022*** **12.8**	***2023*** **10.5**
Preparation time (satisfaction)	*4.1*	*4.23*
Preparation time in days (average)	-	0.6
Median days to request decision	0.6	0.7
Median days to first credit exchange	4.0	1.9
Median days to approved exchange	1.1	1.0
Median days to first resource use	7.1	6.3

Who's Got ACCESS?

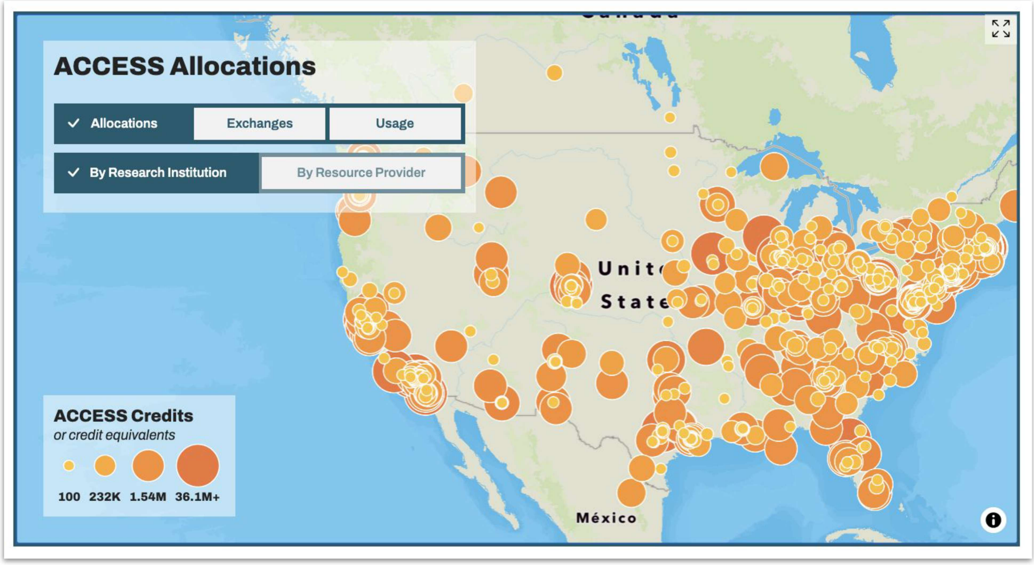
ACCESS Year 2

12,575
new ACCESS accounts

12,624
unique users running jobs

21,454
Support newsletter recipients

57,721
Communications newsletter recipients



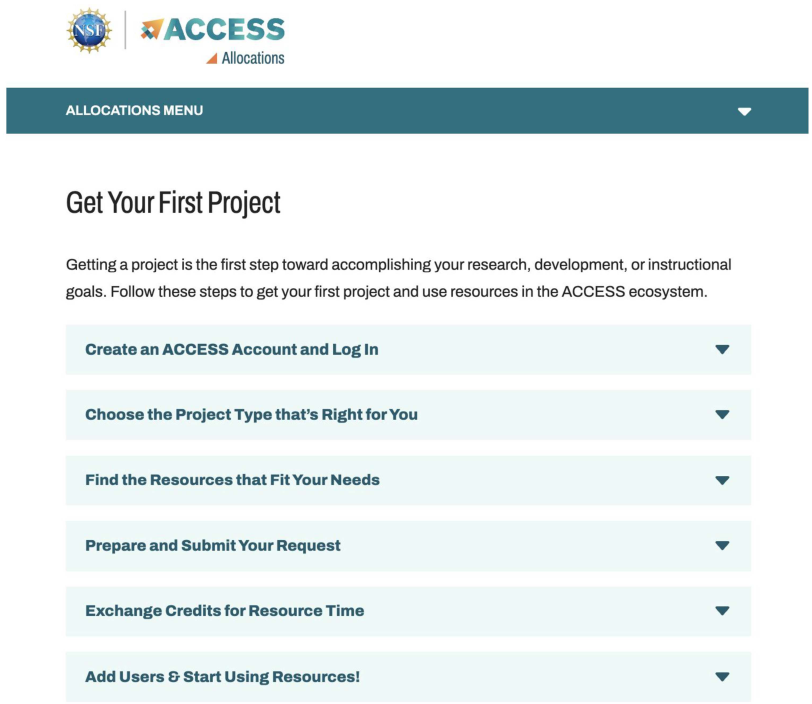
Get Started

To submit a request, you'll need

- Project title & abstract
- Supporting grant info
- A CV and syllabus/training agenda (for instructional activities)

You'll usually have approval for your project within a day.

Projects can last as long as your class or your supporting grant.



allocations.access-ci.org/get-your-first-project

ACCESS to in-depth resource metrics

- Maintain a pulse on ***usage***, ***performance***, and ***behavior*** of NSF-funded ***systems within ACCESS***
- General information about ACCESS cyberinfrastructure use available without signing in
- End users and CIPs can sign in to obtain information on the jobs that they run and to help improve job performance and efficiency.
- Explore data from researchers at your institution at https://xdmod.access-ci.org/

ACCESS for Community Building

The goal for collaboration among the SCIPE, CyberTraining, CSSI and ACCESS programs is to provide tools, information, or resources to enhance the ACCESS ecosystem.

In turn, you can leverage ACCESS services to reach a broader community for your contributions and services.

- Computational Science Support Network (CSSN)
- Affinity Groups
- CI Links
- Ask.CI
- … and more

support.access-ci.org/community/scipe

or contact Alana Romanella <alana.romanella@colorado.edu>

Thank you!

dhart@ucar.edu
access-ci.org

Inquiries — support.access-ci.org/help-ticket

A.3 Metrics for Cybertraining/SCIPE Programs and Outreach

Metrics for Cybertraining/SCIPE Programs and Outreach

Presented By:

Ritu Arora, Wayne State University / Venra Tech

Email: ritu@wayne.edu / ritu@venratech.com

GitHub: https://github.com/ritua2

LinkedIn: https://www.linkedin.com/in/ritu-a-59b58ab/

Twitter: https://twitter.com/ritzaa2

YouTube: https://www.youtube.com/user/ritua2

Some Project Websites: https://icompute.us , https://opuntia.online

Some of the previous related work

- Engaged in upskilling and reskilling CI users and professionals in HPC, cloud computing, AI, and advanced software engineering
- Used team science for building multi-disciplinary collaborations
- Practiced “science of team science” to support evaluation and improvement of team science initiatives
- Researched on development and adoption of “responsible metrics” for assessing the progress, success, and impact of CyberInfrastructure projects including those for workforce development and developed a scoring model for understanding the impacts of the programs and projects
- Provisioned Opuntia – a software infrastructure for facilitating the assessment, discovery, dissemination, and reuse of publicly accessible software and data products – it includes support for sharing metrics of projects and automatically collecting the online mentions/content related to the projects

A reference that includes some examples of metrics for assessing workforce development/outreach projects (1)

Ritu Arora and Sukrit Sondhi. 2024. Towards Developing an Open-Infrastructure for Assessing the Progress, Success, and Impacts of CyberInfrastructure Projects. SN Comput. Sci. 5, 5 (Jun 2024). https://doi.org/10.1007/s42979-024-02961-8

Table 6 PSI metrics for people related projects

Progress	Success	Impact
Number of people who have completed trainings	Training programs or courses meet or exceed all requirements	Number of people who have been placed in new jobs or promoted as a result of trainings or courses
Number of people who have received certifications	Training programs or courses are delivered on time and within budget	Number of people who are employed in CI-related jobs as a result of trainings or courses
Number of instructors engaged	Training programs or courses are effective in increasing the knowledge and skills of participants	Number of people from underrepresented groups who have been hired into CI-related jobs due to trainings or courses
Number of training modules or courses developed or enhanced	The organization has achieved its goal of increasing workforce diversity and inclusiveness	Number of people who are certified in CI-related skills
Number of trainings or courses that were offered	The organization has achieved its goal of increasing the diversity in leadership positions	Number of people who are using the CI skills learned to solve real-world problems
Number of training modules or courses that are digitally accessible	User satisfaction with training programs	Content follows W3C digital accessibility guidelines
Number of email lists or professional organizations or student groups contacted for announcing the availability of trainings or courses		Increased skills of the instructors
Number of activities for broadening participation		Increased public awareness of the technical advancements enabled

A reference that includes some examples of metrics for assessing workforce development/outreach projects (2)

Table 10 People Project- Mapping L1, L2, and L3 Metrics

Project-specific metrics (L1)	Category-specific metrics (L2) People (WD) category selected, category names and PSI dimensions mentioned	General metrics (L3) PSI dimensions mentioned
% of instructors who have completed the trainings on developing inclusive curricula and learning environments	Number of people who have completed trainings (**WD, P**) Number of instructors engaged (**WD, P**) Content follows W3C digital accessibility guidelines (**WD, I**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)
Number of student affinity groups that were contacted to advertise the availability of participant support	Number of email lists or professional organizations or student groups contacted for announcing the availability of trainings or courses (**WD, P**)	Community engagement (**I**)
Number of participants from underrepresented groups that were selected to participate in the summer school	Number of activities for broadening participation (**WD, I**)	Community engagement (**I**)
The number of updates or modifications made to the policies to make the meeting rooms accessible	Number of training modules or courses developed or enhanced (**WD, P**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)
Number of stakeholder meetings that were held to ensure that the project activities themselves are designed in an inclusive manner	Number of activities for broadening participation (**WD, P**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)
% increase in the number of participants from the underrepresented groups in the summer school as compared to the previous years	The organization has achieved its goal of increasing workforce diversity and inclusiveness (**WD, S**)	Outcomes—meeting project goals (**S**)
% increase in the number of accessible meeting rooms at the institution	Number of activities for broadening participation (**WD, P**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)
% increase in the number of instructors who presented inclusive content	Increased skills of the instructors (**WD, I**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)
Are any incidences reported on microaggression?	The organization has achieved its goal of increasing workforce diversity and inclusiveness (**WD, S**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)
Number of participants from underrepresented groups who reported feeling included	The organization has achieved its goal of increasing workforce diversity and inclusiveness (**WD, S**)	Satisfaction—positive feedback from users and stakeholders (**S**)
Positive intention of the participants from the underrepresented groups to apply the knowledge gained at the summer school in research and scholarly activities	Training programs are effective in increasing the knowledge and skills of participants (**WD, S**) Number of people who are using the CI skills learned to solve real-world problem (**WD, I**) The organization has achieved its goal of increasing workforce diversity and inclusiveness (**WD, S**)	Satisfaction—positive feedback from users and stakeholders (**S**)
Implementation of a new policy for developing inclusive curricula (*e.g.*, following W3C accessibility guidelines) at the host institution	Number of training modules or courses that are digitally accessible (**WD, P**) Content follows W3C digital accessibility guidelines (**WD, I**)	Institutional impact—such as, enhancing resources, reputation, skills, and capabilities (**I**)

What is needed for assessing the impacts of the projects and programs?

- Preparing statistically significant (large) and diverse datasets related to project evaluations - by aggregating datasets related to the evaluations of independent programs/projects - so that the findings or results can be generalized by identifying common patterns and developing a set of current best practices
  - A taxonomy of projects
  - A set of standard metrics and the recipes for gathering the data related to the metrics to help in comparing and calibrating the data from different programs/projects belonging to a certain class in the taxonomy
  - A common repository or catalog of evaluation datasets (anonymized) – like Opuntia (https://opuntia.online/) - for sharing quantitative and qualitative data related to the metrics for assessing the progress, success, and impact of the various training, outreach, and other workforce development initiatives
  - A culture of sharing results/knowledge from not so successful efforts such that those can be used as data-points for future improvements - it is important to know “what does/will not work” in addition to knowing “what works”

- Community-driven and semi-automated/automated processes for aggregating the training/education related content developed in the various federally funded projects, and then periodically testing the aggregated content for ensuring continued relevance to the evolving tools and technologies with the help of subject-matter experts and the target audience

- Toolkits for evaluating the workforce development related programs/projects

A.4 Promoting Better Scientific Software

NSF CSSI/CyberTraining/SCIPE PI Meeting
Community experiences and Evolving Needs Panel

Mary Ann Leung
Founder and President
Sustainable Horizons Institute

August 12, 2024

Panel Topics

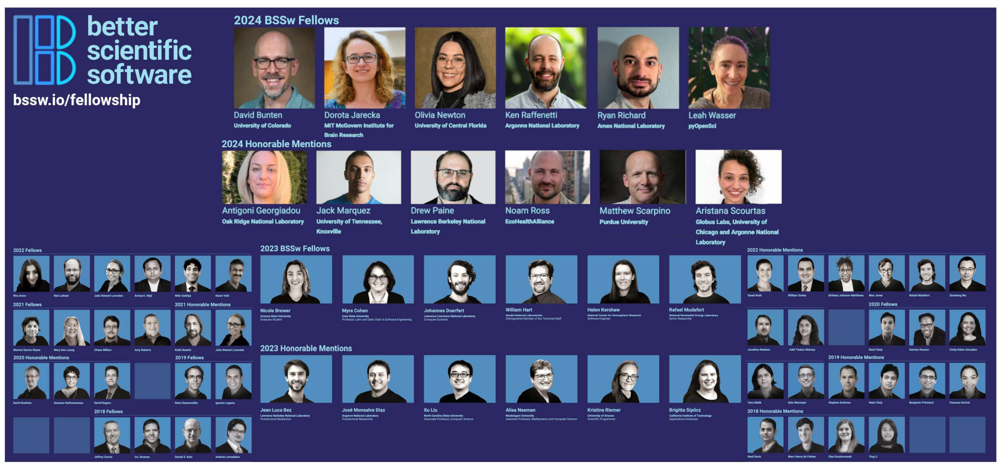
- Metrics for Cybertraining/SCIPE programs and outreach
- Large Institutes - how do they get created?
- Community Building/Broader Impacts:
  - Create broader communities, reach out to MSIs; grow new PIs
- Promoting Better Scientific Software
- New Ways/How to multiply the NSF investment – delivering diverse courses, certificates, programs, different domains; how can they access this?
- How to effectively broaden the science/engineering research workforce, beyond a one size fits all approach

Promoting Better Scientific Software

BSSw Fellowship Program

better scientific software

https://bssw.io/fellowship



BSSw Fellowship Program

better scientific software
https://bssw.io/fellowship

Recognizing leaders and advocates of high-quality scientific software

GOAL

Foster and promote practices, processes, and tools to improve developer productivity and software sustainability of scientific codes

AWARD

- Fellows receive up to $25,000 for an activity promoting better scientific software
- Example activities:
  - organizing a workshop
  - preparing a tutorial
  - creating content to engage the scientific software community

} broad impact that spans beyond a single community/event and results in one or more artifacts that are available beyond the fellowship period of performance

History

2018 through 2024 and beyond

better scientific software
https://bssw.io/fellowship

2018

Launched with DOE support through IDEAS-ECP project.

Awarded 33 fellows + 33 honorable mentions since 2018.

2021

NSF joins to fund and select fellows

Enables more robust approach toward pioneering future cyberinfrastructure leadership

https://bssw.io/blog_posts/nsf-sponsored-2021-bssw-fellows

2024

Current class includes 6 fellows + 6 honorable mentions

2025

Applications open for class of 2025: Aug 15 – Sept 30, 2024

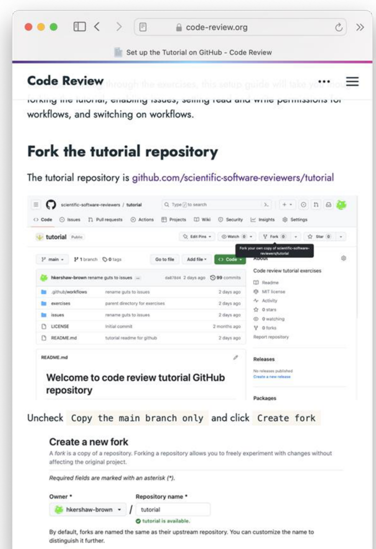
Fellows represent a diverse set of faculty, graduate students, industry professionals, and DOE lab staff

Exemplar BSSw Fellowship Project

Helen Kershaw: 2023 Fellow, NCAR, Software Engineer
Improving code review skills for scientific software developers

DEVELOPING HIGH-QUALITY MATERIALS

Helen's project focused on creating an interactive tutorial that helps developers practice the skill of code review. Typical of most Fellowship projects, she is delivering multiple artifacts, including a GitHub workflow, a stand-alone website, HPC Best Practices webinar (March 2024), and BSSw.io blog article (coming soon) … See *https://code-review.org*



BUILDING ON EXISTING WORK

Although Helen's project is self-contained, it builds on prior work of BSSw Fellows. Helen took inspiration from 2018 Fellow Jeffrey Carver and his project *Improving code quality through modern peer code review.*

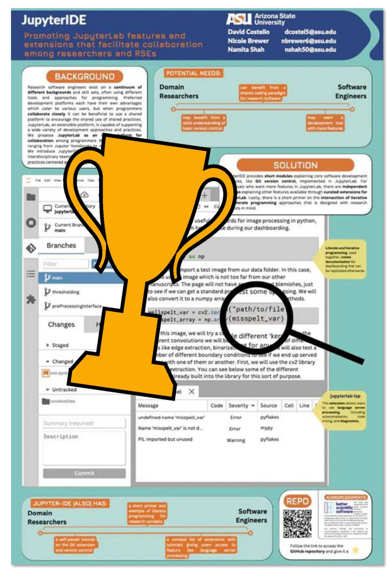
Exemplar BSSw Fellowship Project

Nicole Brewer: 2023 Fellow, Arizona State Univ, Grad Student
Improving accessibility of data and software with scientific web apps

FELLOWSHIP PROJECT

Nicole's project teaches scientists about leveraging Jupyter Notebooks to build web applications. In addition to giving tutorials, she is developing a knowledge repository website. *https://jupyter4science.github.io*

BSSw.io resource (published March 2024):
https://bssw.io/items/jupyter4science-better-practices-for-using-jupyter-notebooks-for-science



BENEFITS FOR A TEAM

With the BSSw Fellowship, Nicole was able to hire two undergraduate interns for summer contributions to her project. One intern, David Costello, won the Best Student Poster Award at US-RSE'23 for his contributions to a set of tutorials about the web-based notebook development environment called JupyterLab.

2024 BSSw Fellowship Class

Work on projects began in March 2024 … https://bssw.io/pages/meet-our-fellows

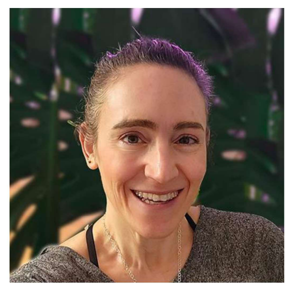
better scientific software
https://bssw.io/fellowship

Dorota Jarecka
MIT McGovern Institute for Brain Research

Best practices for reproducibility and testing in scientific software

Olivia Newton
University of Central Florida

Team learning in scientific software projects



Leah Wasser
pyOpenSci

Essential collaboration skills for contributing to open source software

2024 BSSw Fellowship Class cont.

Work on projects began in March 2024 ... https://bssw.io/pages/meet-our-fellows

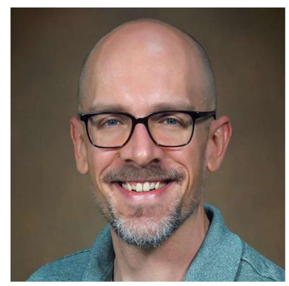
better scientific software
https://bssw.io/fellowship

Ken Raffenetti
Argonne National Laboratory

Guidelines for improving MPI performance

Ryan Richard
Ames National Laboratory

Sustainable scientific software through multi-project CI/CD



David Bunten
University of Colorado Anschutz Medical Campus

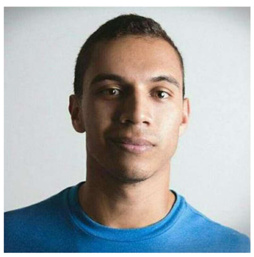
Software gardening almanac: Cultivating sustainable software development in the generative era

2024 BSSw Honorable Mentions

https://bssw.io/pages/meet-our-fellows

better scientific software

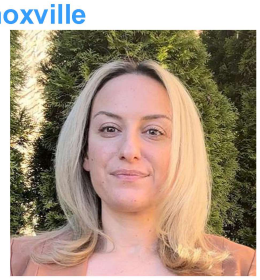
https://bssw.io/fellowship



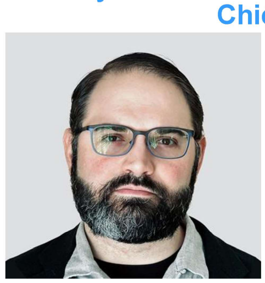
Jack Marquez
University of Tennessee, Knoxville

Matthew Scarpino
Purdue University

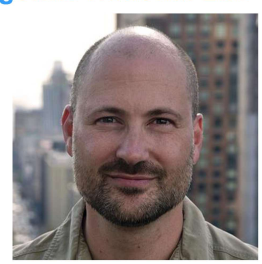
Aristana Scourtas
Globus Labs, University of Chicago and Argonne National Lab



Antigoni Georgiadou
Oak Ridge National Laboratory



Drew Paine
Lawrence Berkeley National Laboratory



Noam Ross
EcoHealthAlliance

Selected Artifacts

better scientific software
https://bssw.io/fellowship

Reaching the broader community

HPC BEST PRACTICES WEBINARS ON YOUTUBE

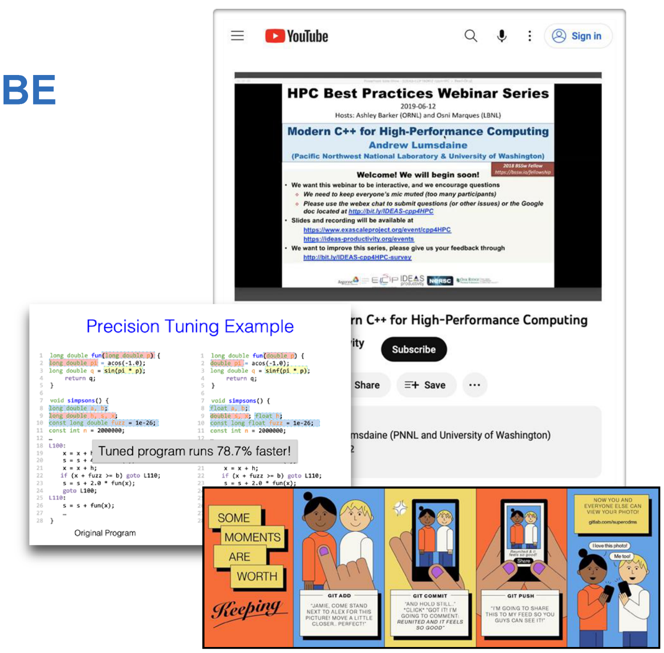
Modern C++ for High-Performance Computing
Andrew Lumsdaine, 2018 Fellow

Scalable Precision Tuning of Numerical Software
Cindy Rubio-Gonzalez, 2020 Fellow

DEDICATED TRAINING WEBSITES

Git User Stories
Amy Roberts, 2021 Fellow

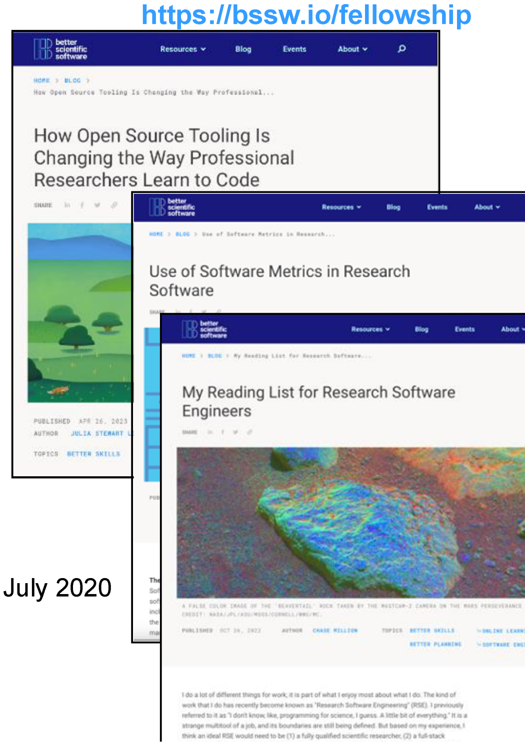
Research Software Development Modules
Kyle Niemeyer, 2019 Fellow



Selected Articles from BSSw.io

Building a community of expertise

https://bssw.io/fellowship



CONTRIBUTIONS FROM THE FELLOWS

Secure Software Programming Practices and Development, Nitin Sukhija, 2022 Fellow
I/O Sleuthing: Digging into Storage Performance, Rob Latham, 2022 Fellow
Containers for Deploying Workflow Systems and Application Codes, Karan Vahi, 2022 Fellow
How Open Source Tooling Is Changing the Way Professional Researchers Learn to Code, Julia Stewart Lowndes, 2022 Fellow
Navigating the Transition of (Climate) Science to the Cloud, Marisol García-Reyes, 2021 Fellow
My Reading List for Research Software Engineers, Chase Million, 2021 Fellow
A Different Way of Looking at Git, Amy Roberts, 2021 Fellow
Use of Software Metrics in Research Software, Nasir Eisty, 2020 Fellow

HIGHLIGHTS OF THE BSSw FELLOWSHIP CLASSES

2022 BSSw Fellows: Projects and Perspectives, July 2023
2021 BSSw Fellows: Projects and Perspectives, July 2022
2020 BSSw Fellows: Projects and Perspectives, July 2021
2019 BSSw Fellows Guide Developers through Each Stage of the Scientific Software Lifecycle, July 2020
2018 BSSw Fellows Tackle Scientific Productivity Challenges, March 2019

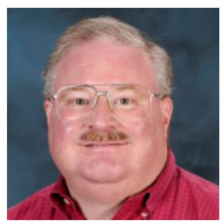
BSSwF Management Partnership

DOE, NSF, and Sustainable Horizons Institute

better scientific software
https://bssw.io/fellowship

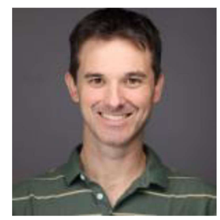
Lois Curfman McInnes
Argonne National Laboratory

Mike Heroux
Sandia National Laboratory

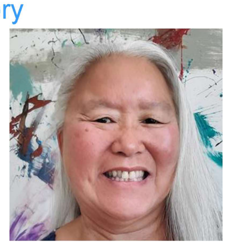


David Bernholdt
Oak Ridge National Laboratory

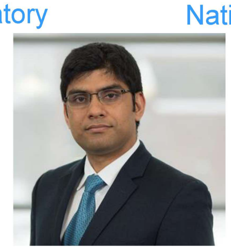
Elsa Gonsiorowski
Lawrence Livermore National Laboratory



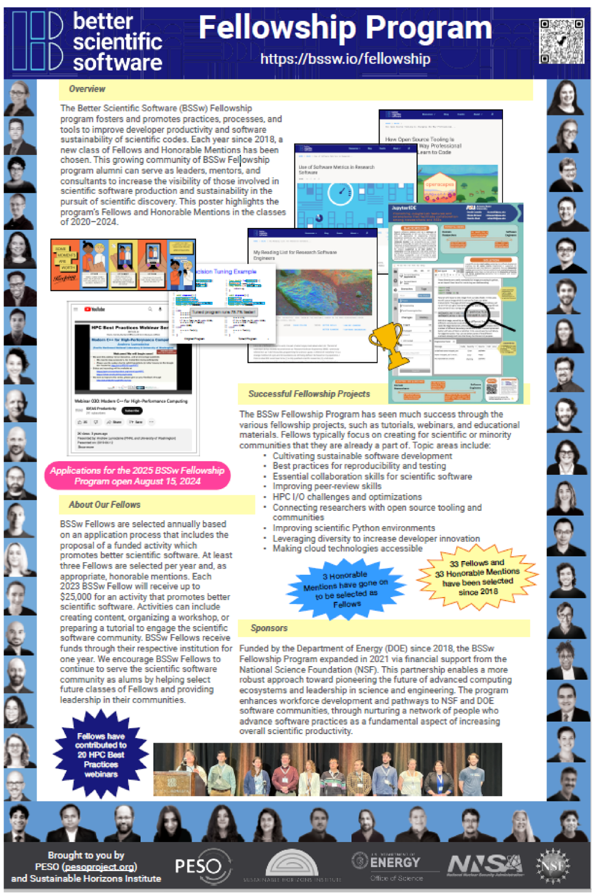
Erik Palmer
Lawrence Berkeley National Laboratory



Mary Ann Leung
Sustainable Horizons Institute



Varun Chandola
National Science Foundation



better scientific software

https://bssw.io/fellowship

BSSw Fellowships

- Building an inclusive community of scientific software professionals and advocates
- Building a collection of resources
  - Webinars
  - Training materials and websites
  - Blogs
  - Etc.

Poster at PEARC24, July 2024

A.5 Democratizing Science through Cyberinfrastructure

Democratizing Science through Cyberinfrastructure

Manish Parashar

Director, Scientific Computing & Imaging (SCI) Institute

Chair in Computational Science and Engineering

Presidential Professor, Kahlert School of Computing

CyberTraining/SCIPE Panel – Community Experiences & Evolving Needs

2024 NSF CSSI/CyberTraining/SCIPE PI Meeting

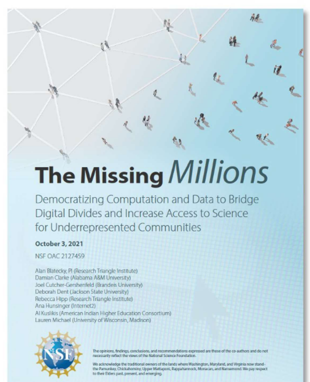
August 12, 2024

Democratizing Science through Cyberinfrastructure

- Broad, fair, and equitable access to advanced is essential to democratizing science in the 21st century

- Significant barriers
  - **Knowledge**: Awareness, discovery, expertise, support
  - **Technical**: Allocation, access, on-ramps
  - **Social**: Awareness of the importance of access to CI, rewards structures

- Realizing an advanced CI ecosystem ***for all***
  - Agile, easily accessible, and scalable **networks of experts** providing embedded expertise and support that is **responsive to local needs.**
  - **Broadly accessible training** targeting the spectrum of CI users and skills.



The Missing *Millions*

Democratizing Computation and Data to Bridge Digital Divides and Increase Access to Science for Underrepresented Communities

October 3, 2021

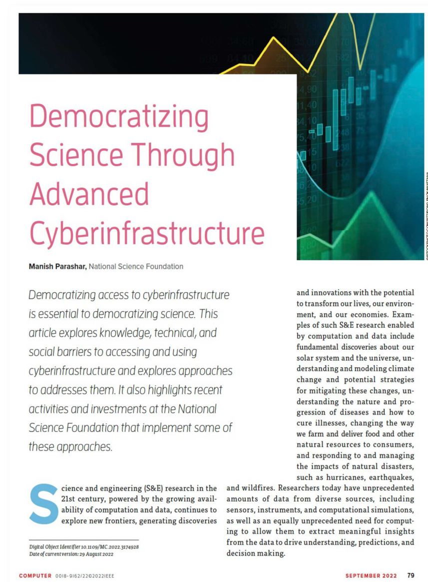
NSF OAC 2127459

Alan Blatecky, PI (Research Triangle Institute)
Damian Clarke (Alabama A&M University)
Joel Cutcher-Gershenfeld (Brandeis University)
Deborah Dent (Jackson State University)
Rebecca Hipp (Research Triangle Institute)
Ana Hunsinger (Internet2)
Al Kuslikis (American Indian Higher Education Consortium)
Lauren Michael (University of Wisconsin, Madison)

The opinions, findings, conclusions, and recommendations expressed are those of the co-authors and do not necessarily reflect the views of the National Science Foundation.

We acknowledge the traditional owners of the lands where Washington, Maryland, and Virginia now stand-the Pamunkey, Chickahominy, Upper Mattaponi, Rappahannock, Monacan, and Nansemond. We pay respect to their Elders past, present, and emerging.

https://www.rti.org/publication/missing-millions/fulltext.pdf



Democratizing Science Through Advanced Cyberinfrastructure

Manish Parashar, National Science Foundation

Democratizing access to cyberinfrastructure is essential to democratizing science. This article explores knowledge, technical, and social barriers to accessing and using cyberinfrastructure and explores approaches to addresses them. It also highlights recent activities and investments at the National Science Foundation that implement some of these approaches.

Science and engineering (S&E) research in the 21st century, powered by the growing availability of computation and data, continues to explore new frontiers, generating discoveries and innovations with the potential to transform our lives, our environment, and our economies. Examples of such S&E research enabled by computation and data include fundamental discoveries about our solar system and the universe, understanding and modeling climate change and potential strategies for mitigating these changes, understanding the nature and progression of diseases and how to cure illnesses, changing the way we farm and deliver food and other natural resources to consumers, and responding to and managing the impacts of natural disasters, such as hurricanes, earthquakes, and wildfires. Researchers today have unprecedented amounts of data from diverse sources, including sensors, instruments, and computational simulations, as well as an equally unprecedented need for computing to allow them to extract meaningful insights from the data to drive understanding, predictions, and decision making.

Digital Object Identifier 10.1109/MC.2022.3174928
Date of current version: 29 August 2022

COMPUTER 0018-9162/22©2022IEEE SEPTEMBER 2022 79

Computer, vol. 55, no. 09, pp. 79-84, 2022. doi: 10.1109/MC.2022.3174928

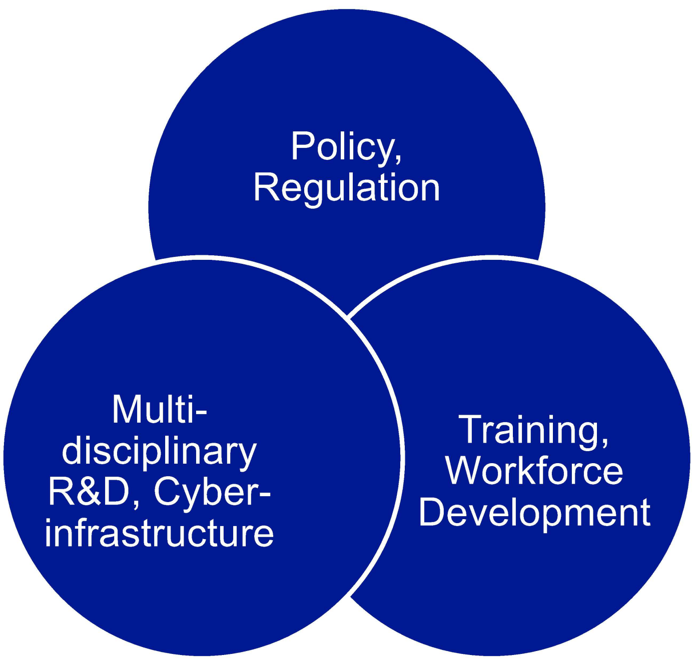
From Open & Equal to Democratized & Equitable Access: Democratizing Access to CI and Science

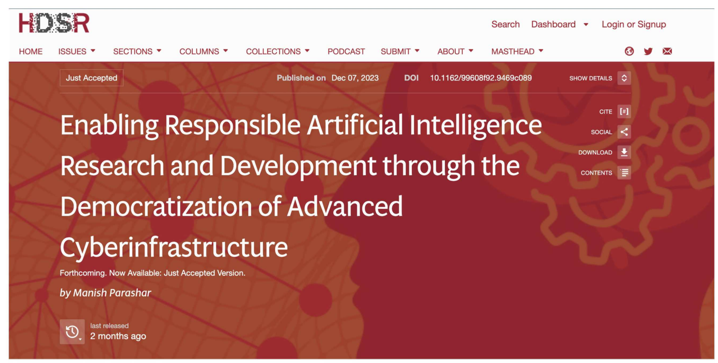
Awareness	• Why is CI important? How is CI relevant to me? What do I need to know about CI? Where do I find the CI that I need?
Ability	• How can I get the skills needed to use CI? Where do I go for help?
Access	• How do I get access to, contribute to, and/or use CI?
Association	• How do I find out who else is using the CI I am using, and how? How do I become part of a community?

From Open & Equal to Democratized & Equitable Access

	OPEN AND EQUAL	DEMOCRATIZED AND EQUITABLE
Allocations	Researchers know of resources, how much needed, write competitive proposals	Flexible allocation mechanisms to meet the needs of the diverse community of researchers
Resources	Researchers have the necessary local infrastructure and system admin support	Local resources as on-ramps to national CI Equitable access using gateways and local access points
Training	Researchers know where to find content, what they need, and have local infrastructure	Readily available pool of local experts providing customized training and on-boarding
User Support	Researchers know what to ask for and whom to ask	Use support and training initiated at the local level based on local needs, with links to a national network of experts
Research	CIP have trusted relations with institution's researchers	Inclusion of locally embedded CIPs into national/regional/topical networks used by researchers to find & develop advanced CI solutions

Democratizing cyberinfrastructure essential for advancing responsible AI





- The quality of research and the pace of innovation are linked to the diversity of the contributions.
- Greater inclusivity in contribution to research and development increases the fairness of the results.

Thank you!

Manish Parashar
Email: manish.parashar@utah.edu
WWW: manishparashar.org / sci.utah.edu / rai.utah.edu

B Participant List, Poster List & DOIs

Overall, there were 286 participants, representing 292 awards funded by CSSI, CyberTraining, OAC Core, CIP, SCIPE CDS & E and related programs. Among the participants, the following NSF Program Directors were present: Sonam Ahluwalia, Amy Apon, Sharmistha Bagchi-Sen, Linkan Bian, Varun Chandola, Sharon Geva, Sheikh Ghafoor, Tom Gulbransen, Marlon Pierce, Plato Smith, Ashok Srinivasan, Wen-wen Tung, Rediet Woldeselassie, Sam Xin, May Yuan.

PARTICIPANT LAST NAME	FIRST NAME	AWARD #	AWARD TITLE	Company
Agar	Joshua	2320600	MRI: Track 2 Development of a Platform for Accessible Data-Intensive Science and Engineering	Drexel University
Akimov	Alexey	1924256	CyberTraining: Pilot: Modeling Excited State Dynamics in Solar Energy Materials	University at Buffalo, SUNY
Akimov	Alexey	1931366	Elements: Libra: The Modular Software for Nonadiabatic and Quantum Dynamics	University at Buffalo, SUNY
Akli	Linda	2400201	CIU: Implementation: Medium: Computational and Data Science Curriculum Exchange Faculty Community of Practice	Texas Advanced Computing Center
Aksamija	Zlatan	2302879	CDSE: Coupled Electro-Thermal Transport in Two-Dimensional Materials and Heterostructures	University of Utah
Amrhein	Dan	2311382	Collaborative Research: Frameworks: A community platform for accelerating observationally-constrained regional oceanographic modeling	NSF NCAR
Andreussi	Oliviero	2321102	Collaborative Research: CyberTraining: Implementation: Medium: Training Users, Developers, and Instructors at the Chemistry/Physics/Materials Science Interface	Boise State University
Arratia	Miguel	2311667	Collaborative Research: Elements: Enabling Particle and Nuclear Physics Discoveries with Neural Deconvolution	UC Riverside
Atherton	Tim	2003820	Elements: Morpho-Cyberinfrastructure for scientists and engineers studying shape change	Tufts University
Aydin	Berkay	2104004	Elements: Spatiotemporal Analysis of Magnetic Polarity Inversion Lines (STEAMPIL)	Georgia State University
Barton	Michael	2103905	Frameworks: Collaborative Research: Integrative Cyberinfrastructure for Next- Generation Modeling Science	Arizona State University/CoMSES.Net
Bass	Steffen	2004571	CSSI: Frameworks: X-Ion Collisions with a Statistically and Computationally Advanced Program Envelope (X-SCAPE)	Duke University
Bhasi	Vivek	1931531		Penn State
Bhatia	Sajal	2017371	CyberTraining: Implementation: Small: Using Problem-Based Learning for Vocational Training in Cyberinfrastructure Security at Community Colleges	Sacred Heart University
Bhowmick	Sanjukta	2104076	Collaborative Research: Framework Implementations: CSSI: CANDY: Cyberinfrastructure for Accelerating Innovation in Network Dynamics	University of North Texas
Bockelman	Brian	2103963	CSSI Elements: EWMS - Event Workflow Management Service	Morgridge Institute for Research
Bockelman	Brian	2209645	Elements: Kingfisher: Storage Management for Data Federations	Morgridge Institute for Research
Bouteiller	Aurelien	1931384	CSSI: EPEXA Ecosystem for Programming and Executing eXtreme-scale Applications	The University of Tennessee
Brandt	Kevin	2227656	RCN:CIP: A Connect.CI-based Community-Wide Mentorship Network (CCMnet) for the Advancement of Science and Engineering Research and Education	South Dakota State University
Brinson	Cate	1835677	Nanocomposites to Metamaterials: A Knowledge Graph Framework	Duke University
Buitrago	Paola	2320991	Collaborative Research: CyberTraining: Pilot: Building a strong community of computational researchers empowered in the use of novel cutting-edge technologies	Carnegie Mellon University
Candan	K. Selcuk	2311716	Elements: CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration	Arizona State University
Carifio	Jonathan	2209623, 2209624, 2209625	Collaborative Research: Elements: Enriching Scholarly Communication with Augmented Reality	Center for Astrophysics | HarvardSmithsonian
Carr	Steve	2320951	CyberTraining: Implementation: Small: Promoting AI Readiness for Machine-Assisted Secure Data Analysis (PAIR4MASDA)	Western Michigan University
Carver	Jeff	2017259	Collaborative Research: CyberTraining: Implementation: Small: INnovative Training Enabled by a Research Software Engineering Community of Trainers (INTERSECT)	University of Alabama
Carver	Jeff	2209821	Collaborative Research: Frameworks for Intelligent Adaptive Experimentation: Enhancing and Tailoring Digital Education	University of Alabama
Casanova	Henri	2103489	Collaborative Research: Elements: Simulation-driven Evaluation of Cyberinfrastructure Systems	University of Hawaii
Chakraborty	Ankit	2103524	Elements:Software A Scalable Open-Source hp-Adaptive FE Software for Complex Multiphysics Applications	The University of Texas at Austin
Chakravorty	Dhruva	2320992	CyberTraining: Pilot: Building a strong community of computational researchers empowered in the use of novel cutting-edge technologies	Texas AM University
Chard	Kyle	2004894	Collaborative Research: Frameworks: funcX: A Function Execution Service for Portability and Performance	University of Chicago
Chard	Kyle	2104008	Frameworks: Collaborative Research: ChronoLog: A High-Performance Storage Infrastructure for Activity and Log Workloads	University of Chicago
Chard	Kyle	2209919	Collaborative Research: Sustainability: A Community-Centered Approach for Supporting and Sustaining Parsl	University of Chicago
Chaudhary	Vipin	2320952	Collaborative Research: SCIPE: Interdisciplinary Research Support Community for Artificial Intelligence and Data Sciences	Case Western Reserve University
Chen	Weicong	2321123	Cross-Layer Training of High-Performance Deep Learning Technologies and Applications for Research Workforce Development in Central Valley	University of California, Merced
Chilton	Lydia	2103794	Elements: Open-Source Cyberinfrastructure as a Decision Engine for Socioeconomic Disaster Risk	Columbia University
Choi	Eunseo	2104002	Elements: Developing an integrated modeling platform for tectonics, earthquake cycles and surface processes	The University of Memphis
Chourasia	Amit	2231406	CI CoE: SGX3 - A Center of Excellence to Extend Access, Expand the Community, and Exemplify Good Practices for CI Through Science Gateways	UCSD
Chourasia	Amit	2311206, 2311207, 2311208	Collaborative Research: Frameworks: Quakeworx - An extensible software framework for earthquake simulations	UCSD
Cleveland	Sean	1931575	Collaborative Proposal: Frameworks: Project Tapis: Next Generation Software for Distributed Research	University of Hawaii - System
Cleveland	Sean	2118222	CyberTraining: Implementation: Medium: Cyberinfrastructure Training to Advance Environmental Science	University of Hawaii - System
Closser	Kristi	2417553	CyberTraining: Implementation: Medium: ChemCompute: Computational Chemistry for Undergraduate Education and Research	California State University, Fresno
Codoni	David	2310909	Collaborative Research: Frameworks: A multi-fidelity computational framework for vascular mechanobiology in SimVascular	Stanford University
Coles	Victoria	2321008	Collaborative Research: SCIPE: Enhancing the Transdisciplinary Research Ecosystem for Earth and Environmental Science with Dedicated Cyber Infrastructure Professionals	University of Maryland Center for Environmental Science
Combs	Jane	2320953	Collaborative Research: SCIPE: Interdisciplinary Research Support Community for Artificial Intelligence and Data Sciences	University of Cincinnati
Corbin	Nicole	2319979	CyberTraining: Pilot: Machine Learning Foundations and Applications in the Earth Systems Sciences	University Corporation for Atmospheric Research (UCAR)
Crawford	Daniel	2136142	S2I2: Impl: The Molecular Sciences Software Institute	Virginia Tech
Crawford	Daniel	2321044	CyberTraining: Implementation: Medium: Transforming the Molecular Science Research Workforce through Integration of Programming in University Curricula	Virginia Tech
Crichigno	Jorge	2118311	CyberTraining: Implementation: Small: Cybertraining on P4 Programmable Devices using an Online Scalable Platform with Physical and Virtual Switches and Real Protocol Stacks	University of South Carolina
Crichigno	Jorge	2403360	OAC Core: Enhancing Network Security by Implementing an ML Malware Detection and Classification Scheme in P4 Programmable Data Planes and SmartNICs	University of South Carolina
Crosby	Lonnie	2230106	Collaborative Research: CyberTraining: CIP: A Cross-Institutional Research Engagement Network for CI Facilitators	University of Tennessee, Knoxville
CROZIER	PETER	2104105	Elements: Collaborative Research: Community-driven Environment of AI-powered Noise Reduction Services for Materials Discovery from Electron Microscopy Data	Arizona State University
Cui	Hantao	2319895	Collaborative Research: CyberTraining: Pilot: PowerCyber: Computational Training for Power Engineering Researchers	North Carolina State University
Cui	Suxia	2321111	Collaborative Research: Train the Trainers as Next Generation Leaders in Data Science for Cybersecurity for Underrepresented Communities	Prairie View AM University
D'Orazio	Vito	2311142	Frameworks: Infrastructure for Political and Social Event Data Using Machine Learning	West Virginia University
Davidson	Drew	2311819	Collaborative Research: Frameworks: Automated Quality Assurance and Quality Control for the StraboSpot Geologic Information System and Observational Data	University of Kansas
Deelman	Ewa	2103508	Collaborative Research: Elements: Simulation-driven Evaluation of Cyberinfrastructure Systems	University of Southern California
Deelman	Ewa	2320375	Collaborative Research: CyberTraining: Implementation: Medium: CyberInfrastructure Training and Education for Synchrotron X-Ray Science (X-CITE)	University of Southern California
Denolle	Marine	2103494	Seismic COmputational Platform for Empowering Discovery	U Washington
Denolle	Marine	2103621	Seismic Computational Platform for Empowering Discovery	U Washington
Denolle	Marine	2103701	Seismic COmputational Platform for Empowering Discovery	U Washington
Denolle	Marine	2103741	Seismic COmputational Platform for Empowering Discovery	U Washington
Denolle	Marine	2104052	Seismic COmputational Platform for Empowering Discovery	U Washington
Deppe	Nils	2209655	Collaborative Research: Elements: A task-based code for multiphysics problems in astrophysics at exascale	Cornell University
DePrince	Eugene	2103705	Collaborative Proposal: Frameworks: Sustainable Open-Source Quantum Dynamics and Spectroscopy Software	Florida State University
Di	Sheng	2104023, 2247080	CSSI: Elements: ROCCI: Integrated Cyberinfrastructure for In Situ Lossy Compression Optimization Based on Post Hoc Analysis Requirements	University of Chicago
Dong	Wenqian	2417849	Collaborative Research: CyberTraining: Implementation: Medium: EcoTern: Pioneering a CI Workforce for Sustainable and Transdisciplinary Environmental Science Research	Oregon State University
Dunbar	Oliver	1835576	Data-Driven Earth System Modeling	California Institute of Technology
Dunbar	Oliver	1835860	Data-Driven Earth System Modeling	California Institute of Technology

PARTICIPANT LAST NAME	FIRST NAME	AWARD #	AWARD TITLE	Company
Eisma	Jessica	2230054	CyberTraining: Pilot: Justice in Data: An intensive, mentored online bootcamp developing FAIR data competencies in undergraduate researchers in the water and energy sectors	University of Texas at Arlington
Esfarjani	Keivan	2103989	ELEMENTS: Anharmonic formalism and codes to calculate thermal transport and phase change from first-principles calculations	University of Virginia
Evans	John	2104106	Collaborative Research: Elements: EXHUME: Extraction for High-Order Unfitted Finite Element Methods	University of Colorado Boulder
Figueiredo	Renato	2311123	Collaborative Research: Elements: FaaSr: Enabling Cloud-native Event-driven Function-as-a-Service Computing Workflows in R	University of Florida
Figueiredo	Renato	2311124	Collaborative Research: Elements: FaaSr: Enabling Cloud-native Event-driven Function-as-a-Service Computing Workflows in R	University of Florida
Foster	Ian	2209892	Frameworks: Garden: A FAIR Framework for Publishing and Applying AI Models for Translational Research in Science, Engineering, Education, and Industry	University of Chicago and Argonne National Laboratory
Ganapati	Sukumar	1924154	CyberTraining: Implementation: Medium: Advanced Cyber Infrastructure Training in Policy Informatics	Florida International University
Gates	Mark	2004541	Collaborative Research: Frameworks: Basic ALgebra LIbraries for Sustainable Technology with Interdisciplinary Collaboration (BALLISTIC)	University of Tennessee
Georgoulis	Manolis	2104004	Elements: Spatiotemporal Analysis of Magnetic Polarity Inversion Lines (STEAMPIL)	Johns Hopkins Applied Physics Laboratory
Gerber	Edwin	2004572	Collaborative Research: Framework: Improving the Understanding and Representation of Atmospheric Gravity Waves using High-Resolution Observations and Machine Learning	New York University
Gesing	Sandra	2231406	CI CoE: SGX3 - A Center of Excellence to Extend Access, Expand the Community, and Exemplify Good Practices for CI Through Science Gateways	US Research Software Engineer Association
Giustino	Feliciano	2103991	Frameworks: An Interoperable Software Ecosystem for Many-Body Electronic Structure Calculations	UT Austin
Griffith	Boyce	1931516	Multiphase Fluid-Structure Interaction Software Infrastructure to Enable Applications in Medicine, Biology, and Engineering	University of North Carolina at Chapel Hill
Guan	Qiang	2230111	CyberTraining: Implementation: Small: Interactive and Integrated Training for Quantum Application Developers across Platforms	Kent State University
Gull	Emanuel	2310582	Elements: Embedding Framework for Quantum Many-Body Simulations	University of Michigan
Haas	Roland	2004879	Collaborative Research: Frameworks: The Einstein Toolkit ecosystem: Enabling fundamental research in the era of multi-messenger astrophysics	University of Illinois
Haas	Roland	2310548	Elements: An initial value solver for the era of multi-messenger astrophysics	University of Illinois
Hamilton	Kathryn	2311928	Frameworks: An Advanced Cyberinfrastructure for Atomic, Molecular, and Optical Science (AMOS): Democratizing AMOS for Research and Education	University of Colorado Denver
He	Suining	2118102	CyberTraining: Pilot: Cyberinfrastructure Training in Computer Science and Geoscience	University of Connecticut
He	Xubin	2311758	Collaborative Research: Elements: ProDM: Developing A Unified Progressive Data Management Library for Exascale Computational Science	Temple University
Heffernan	Neil	1931523	Title:Collaborative Research: Frameworks: Cyber Infrastructure for Shared Algorithmic and Experimental Research in Online Learning	Worcester Polytechnic Institute
Heinz	Hendrik	1931587	CSSI Framework: Cyberloop for Accelerated Bionanomaterials Design	University of Colorado Boulder
Heyden	Matthias	2311372	Elements: Streaming Molecular Dynamics Simulation Trajectories for Direct Analysis: Applications to Sub-Picosecond Dynamics in Microsecond Simulations	Arizona State University
Hillery	Elizabett	2321090	Collaborative Research: SCIPE: CyberInfrastructure Professionals InnoVating and brOadening the adoption of advanced Technologies (CI PIVOT)	Purdue University
Howard	Michael	2310724	Collaborative Research: Elements: Multiparticle collision dynamics simulations of mesoscale hydrodynamic interactions in complex soft materials and environments	Auburn University
Howard	Michael	2310725	Collaborative Research: Elements: Multiparticle collision dynamics simulations of mesoscale hydrodynamic interactions in complex soft materials and environments	Auburn University
Hu	Ming	2311202	Collaborative Research: Elements: Phonon Database Generation, Analysis, and Visualization for Data Driven Materials Discovery	University of South Carolina
Hudak	David	1835725	Frameworks: Software NSCI-Open OnDemand 2.0: Advancing Accessibility and Scalability for Computational Science through Leveraged Software Cyberinfrastructure	Ohio Supercomputer Center
Hudak	David	2320954	Collaborative Research: SCIPE: Interdisciplinary Research Support Community for Artificial Intelligence and Data Sciences	Ohio Supercomputer Center
Hutton	Eric	2104102	Collaborative Research: Frameworks: OpenEarthscape - Transformative Cyberinfrastructure for Modeling and Simulation in the Earth-Surface Science Communities	University of Colorado
Jafarzadeh	Ashkan	2311527	Enabling High Precision Citizen Science Photometry from Wide Field Color Images	University of Texas at Dallas
Jawed	Khalid	2209782	Collaborative Research: Elements: Discrete Simulation of Flexible Structures and Soft Robots	University of California, Los Angeles
Jee	Kangkook	2321117	CyberTraining:Pilot:CyberTraining for Space CI in Low Earth Orbit (LEO)	The University of Texas at Dallas
Jiang	Weiwen	2320957	CyberTraining: Pilot: Quantum Research Workforce Development on End-to-End Quantum Systems Integration	George Mason University
Kalyanam	Rajesh	2311762	Elements: Science-i Cyberinfrastructure for Forest Ecosystem Research	Purdue University
Kandasamy	Nagarajan	2209745	Elements: Software Infrastructure for Programming and Architectural Exploration of Neuromorphic Computing Systems	Drexel University
Karagiorgi	Georgia	2209917	Elements: RAD Discoveries for Fundamental Physics	Columbia University
Katsavounidis	Erik	1931469	Collaborative Research: Frameworks: Machine learning and FPGA computing for real-time applications in big-data physics experiments	MIT
Katz	Daniel S.	2004932	Collaborative Research: Frameworks: funcX: A Function Execution Service for Portability and Performance	University of Illinois
Katz	Daniel S.	2209920	Collaborative Research: Sustainability: A Community-Centered Approach for Supporting and Sustaining Parsl	University of Illinois
Kerzendorf	Wolfgang	2311323	Elements: The TARDIS radiative transfer framework - A modeling toolkit for transients	Michigan State University
Kim	Hyesoon	2103951	Elements:Open-source hardware and software evaluation system for UAV	Georgia Tech
Kline Struhl	Melissa	2209756	CSSI Frameworks: Cyberinfrastructure for Remote Data Collection with Children	MIT
Knepley	Matthew	1931524	Collaborative Research: Frameworks: Multiphase Fluid-Structure Interaction Software Infrastructure to Enable Applications in Medicine, Biology, and Engineering	University at Buffalo
Knepper	Richard	2209711	Frameworks: Large Scale Atmospheric Research Using an Integrated WRF Modeling, Visualization, and Verification Container Framework (I-WRF)	Cornell University
Knepper	Richard	2320977	CyberTraining: Pilot: HPC ED: Building a Federated Repository and Increasing Access through Cybertraining	Cornell University
Knight	Christopher	2311260	Frameworks: Data-Driven Software Infrastructure for Next-Generation Molecular Simulations	University of Chicago/Argonne
Kumar	Ratnesh	2004766	Elements: Agricultural Cyber-infrastructure support for Field and Grid Modeling, and Runtime Decision-Making	Iowa State Univ of Science and Technology
Kumar	Krishna	2103937	Elements: Cognitasium - Enabling Data-Driven Discoveries in Natural Hazards Engineering	University of Texas at Austin
Kumar	Krishna	2229702	POSE: Phase I: Tuitus - A sustainable, inclusive, open ecosystem for Natural Hazards Engineering	University of Texas at Austin
Kumar	Krishna	2321040	SCIPE: Chishiki.ai: A sustainable, diverse, and integrated CIP community for Artificial Intelligence in Civil and Environmental Engineering	University of Texas at Austin
Lahoti	Radha Manoj	2209782	Collaborative Research: Elements: Discrete Simulation of Flexible Structures and Soft Robots	University of California Los Angeles
Lange	David	2311471	CLAD Enabling Differentiable Programming in Science	Princeton University
Lejeune	Emma	2310771	Elements: Curating and Disseminating Solid Mechanics Based Benchmark Datasets	Boston University
Lele	Sanjiva	2103509	Adaptive Multi-resolution High-order Solver for Multiphase Compressible Flows on Heterogeneous Platforms	Stanford University
Lemson	Gerard	2311791	Open SciServer: A Sustainable Data-Driven Science Platform	the Johns Hopkins University
Lewis	Thu-Mai	2209630	Collaborative Research: Elements: TRAnsparency CErtified (TRACE): Trusting Computational Research Without Repeating It	University of North Carolina at Chapel Hill
Li	Dong	2104116	Collaborative Research: Elements: SciMem: Enabling High Performance Multi-Scale Simulation on Big Memory Platforms	University of California, Merced
Li	Wenwen	2230034	Collaborative Research: CyberTraining: Implementation: Medium: Cyber2A: CyberTraining on AI-driven Analytics for Next Generation Arctic Scientists	Arizona State University
Liang	Xu	2209833	Collaborative Research: Frameworks: Building a collaboration infrastructure: CyberWater sustainable data/model integration framework	University of Pittsburgh
Liang	Xin	2311756	Collaborative Research: Elements: ProDM: Developing A Unified Progressive Data Management Library for Exascale Computational Science	University of Kentucky
Liang	Xin	2330364	Collaborative Research: CyberTraining: Pilot: Research Workforce Development for Deep Learning Systems in Advanced GPU Cyberinfrastructure	University of Kentucky
Lindner	Peggy	2321110	Collaborative Research: CyberTraining: Implementation: Small: Train the Trainers as Next Generation Leaders in Data Science for Cybersecurity for Underrepresented Communities	University of Houston
Lindzey	Laura	2209726	Elements: Making Ice Penetrating Radar More Accessible: A tool for finding, downloading and visualizing georeferenced radargrams within the QGIS ecosystem	University of Washington Applied Physics Laboratory
Liu	Dunyu	2209807	Collaborative Research: Elements: Monitoring Earth Surface Deformation with the Next Generation of InSAR Satellites: GMTSAR	University of Texas
Liu	Qianqian	2230046	Collaborative Research: CyberTraining: Pilot: A Cybertraining Program to Advance Data Acquisition, Processing, and Machine Learning-based	University of North Carolina Wilmington
Logan	Luke	2313154	OAC Core: LABIOS: Storage Acceleration via Data Labeling and Asynchronous I/O	Illinois Institute of Technology
Lou	Helen	2321055	CyberTraining: Pilot: Interdisciplinary Cybersecurity Education to Support Critical Energy and Chemical Infrastructure	Lamar University
Lowe-Power	Jason	2311888	Collaborative Research: Frameworks: Advancing Computer Hardware and Systems' Research Capability, Reproducibility, and Sustainability with the gem5 Simulator Ecosystem	UC, Davis
Lu	Yung Hsiang	2104709	CDSE: Collaborative: Cyber Infrastructure to Enable Computer Vision Applications at the Edge Using Automated Contextual Analysis	Purdue University
Lu	Yung Hsiang	2107230	Collaborative Research: OAC Core: Advancing Low-Power Computer Vision at the Edge	Purdue University
Lu	Xiaoyi	2321123	CyberTraining: Pilot: Cross-Layer Training of High-Performance Deep Learning Technologies and Applications for Research Workforce Development in Central Valley	University of California, Merced
Lu	Guoyu	2334690	Elements: A Deep Neural Network-based Drone (UAS) Sensing System for 3D Crop Structure Assessment	University of Georgia
Lv	Qin "Christine"	2311843	Elements: Development of /cyberinfrastructure to establish a scalable application of self-supervised machine learning for over a decade of NOAA's water column sonar data	University of Colorado Boulder
Lynch	Ben	2320769	Cyber Training Pilot: Breaking the Compute Barrier, Upskilling Agri-Food Researchers to Utilize HPC Resources	University of Minnesota

PARTICIPANT LAST NAME	FIRST NAME	AWARD #	AWARD TITLE	Company
Machi	Dustin	1916805	CINES: A Scalable Cyberinfrastructure for Sustained Innovation in Network Engineering and Science	University of Virginia
Maffeo	Christopher	2311550	Elements: Enabling multi-resolution simulations at the interface of biology and nanotechnology with ARBD	University of Illinois at Urbana-Champaign
Mandal	Subhasish	2311558	Collaborative Research: Elements: Building an open source DFT+eDMFT database for quantum materials	West Virginia University
Mandal	Anirban	2320373	Collaborative Research: CyberTraining: Implementation: Medium: CyberInfrastructure Training and Education for Synchrotron X-Ray Science (X-CITE)	UNC Chapel Hill
Manning	T. Andrew	2103680	Frameworks: MUSES, Modular Unified Solver of the Equation of State	National Center for Supercomputing Applications, University of Illinois Urbana-Champaign
Marru	Suresh	2209872	Collaborative Research: Frameworks: Cybershuttle: An end-to-end Cyberinfrastructure Continuum to accelerate Discovery in Science and Engineering	Georgia Tech
Martiniani	Stefano	2311632	GOALI: Frameworks: At-Scale Heterogeneous Data based Adaptive Development Platform for Machine-Learning Models for Material and Chemical Discovery	New York University
May	Ryan	2103682	Elements: Scaling MetPy to Big Data Workflows in Meteorology and Climate Science	UCAR
McCully	Curtis	2311355	Frameworks: SCiMMA: Real-time Orchestration of Multi-Messenger Astrophysical Observations	Las Cumbres Observatory
McHenry	Kenton	2230035	Collaborative Research: CyberTraining: Implementation: Medium: Cyber2A: CyberTraining on AI-driven Analytics for Next Generation Arctic Scientists	University of Illinois
Meneveau	Charles	2103874	Frameworks: Advanced Cyberinfrastructure for Sustainable Community Usage of Big Data from Numerical Fluid Dynamics Simulations	Johns Hopkins University
Menzies	Tim	1931425	Elements: Can Empirical SE be Adapted to Computational Science?	Nc State University
Merz	Kenneth	2209717	Collaborative Research: Frameworks: Interoperable High-Performance Classical, Machine Learning and Quantum Free Energy Methods in AMBER	MSU
Mills	William	2311919	Sustainability: Long-term Deployment Sustainability Strategy for Argovis	University of Colorado
Moore	Shirley	2311708	Collaborative Research: Frameworks: Scalable Performance and Accuracy analysis for Distributed and Extreme-scale systems (SPADE)	University of Texas at El Paso
Moreno-Centeno	Erick	1835499	Elements: Software: Roundoff-Error-Free Algorithms for Large-Scale, Sparse Systems of Linear Equations and Optimization	Texas AM University
Morgan	Dane	1931306	Collaborative Research: Framework: Machine Learning Materials Innovation Infrastructure	Univ of Wisconsin - Madison
Morgan	Dane	2017072	The Informatics Skunkworks Program for Undergraduate Research at the Interface of Data Science and Materials Science and Engineering	Univ of Wisconsin - Madison
Morse	David	2103627	Elements: Open-sources tools for block polymer phase behavior	University of Minnesota
Nadeem	Tamer	2212424	OAC Core: MedKnights - Towards Secure and Flexible Medical IoT (IoMT) Infrastructure using Generative Adversarial Networks	Virginia Commonwealth University
Nakano	Aiichiro	2118061	Collaborative Research: CyberTraining: Implementation: Medium: Cyber Training on Materials Genome Innovation for Computational Software (CyberMAGICS)	University of Southern California
Narayanan	Sri Hari Krishna	2104068	Collaborative Research: Frameworks: Convergence of Bayesian inverse methods and scientific machine learning in Earth system models through universal differentiable programming	Argonne National Laboratory, UChicago CASE
Nawab	Faisal	2321121	Collaborative Research: CyberTraining: Pilot: Cybertraining to Develop FAIR Data Competencies for Bioengineering Students	University of California, Irvine
Neeman	Henry	2118193	CyberTraining: Pilot: A Professional Development and Certification Program for Cyberinfrastructure Facilitators	University of Oklahoma
Negrut	Dan	2209791, 2209792, 2209793, 2209794, 2209795	Collaborative Research: Frameworks: Simulating Autonomous Agents and the Human-Autonomous Agent Interaction	University of Wisconsin-Madison
Ni	Zhen	1949921	Collaborative Research: CyberTraining: Implementation: Small: Multi-disciplinary Training of Learning, Optimization and Communications for Next Generation Power Engineers	Florida Atlantic University
Ni	Zhen	2320972	Collaborative Research: Implementation: Medium: Secure, Resilient Cyber-Physical Energy System Workforce Pathways via Data-Centric, Hardware-in-the-Loop Training	Florida Atlantic University
Nicolae	Bogdan	2411386	VLCC-States: Versioned Lineage-Driven Checkpointing of Composable States	University of Chicago
Niu	Wei	2403090	Collaborative Research: OAC Core: CropDL - Scheduling and Checkpoint/Restart Support for Deep Learning Applications on HPC Clusters	University of Georgia
North	Chris	2004014, 2003800, 2003387	Collaborative Research: CSSI Frameworks: SAGE3: Smart Amplified Group Environment for Harnessing the Data Revolution	Virginia Tech
Olaya	Paula	2103845	Collaborative Research: Elements: SENSORY: Software Ecosystem for kNowledge diScOveRY - a data-driven framework for soil moisture applications	University of Tennessee, Knoxville
Ou	Ge	2004658	Elements: Open Access Data Generation Engine for Bulk Power System under Extreme Windstorms	University of FLorida
Pallickara	Shrideep	1931363	Frameworks: Collaborative Proposal: Software Infrastructure for Transformative Urban Sustainability Research.	Colorado State University
Panda	Dhabaleswar K (DK)	2311830	Collaborative Research: Frameworks: Performance engineering scientific applications with MVAPICH and TAU using emerging communication primitives	The Ohio State University
Pantano-Rubino	Carlos	2320943	Pilot: Fostering Computational Excellence (FOCEX): Addressing the Disconnect between Advanced CyberInfrastructure and Educational Preparedness	University of Southern California
Park	Seung-Jong	2403399	Collaborative Research: OAC: Core: Harvesting Idle Resources Safely and Timely for Large-scale AI Applications in High-Performance Computing Systems	Missouri University of Science and Technology
Pascucci	Valerio	2138811	OAC: Piloting the National Science Data Fabric: A Platform Agnostic Testbed for Democratizing Data Delivery	CEDMAV, University of Utah
Patel	Vivak	2309445	Collaborative Research: Elements: A Cyberlaboratory for Randomized Numerical Linear Algebra	University of Wisconsin, Madison
Patra	Abani	2004826 - 1928406	GLAcier Feature Tracking testkit (GLAFT): a statistically and physically based framework for evaluating glacier velocity products derived from optical satellite image feature tracking	Tufts University
Patra	Abani	2004826 - 2004302	Firn aquifer water discharges into crevasses across Southeast Greenland	Tufts University
Patra	Abani	2004826 2125974 2004302	GHub : Building a glaciology gateway to unify a community	Tufts University
Peng	Shiyu	2209262	Elements: The PERTURBO Package: A Community Code for Electron Interactions and Dynamics in Materials	California Institute of Technology
Philippov	Alexander	2311800	Elements: Entity: Radiative General-Relativistic Particle-in-cell Toolkit for Extreme Plasma Astrophysics	University of Maryland
Phillips	Daniel	2004601	Frameworks: Bayesian Analysis of Nuclear Dynamics	Ohio University
Piper	Mark	2103878	Frameworks: Collaborative Research: Integrative Cyberinfrastructure for Next-Generation Modeling Science	University of Colorado Boulder
Piper	Mark	2118171	Collaborative Research: CyberTraining: Pilot: A Cybertraining Program to Advance Knowledge and Equity in the Geosciences	University of Colorado Boulder
Pivarski	Jim	2103945	Framework: Awkward Arrays - Accelerating scientific data analysis on irregularly shaped data	Princeton University
Puri	Satish	2344585	Collaborative Research: OAC: Approximate Nearest Neighbor Similarity Search for Large Polygonal and Trajectory Datasets	Missouri University of Science and Technology
Purwanto	Wirawan	2320998	T3-CIDERS: A Train-the-Trainer Approach to Fostering CI- and Data-Enabled Research in Cybersecurity	Old Dominion University
Quinn	Daven	2311091	Collaborative Research: Elements: Linking geochemical proxy records to crustal stratigraphic context via community-interactive cyberinfrastructure	University of Wisconsin Madison
Quinn	Daven	2311092	Collaborative Research: Elements: Linking geochemical proxy records to crustal stratigraphic context via community-interactive cyberinfrastructure	University of Wisconsin, Madison
Rafique	M. Mustafa	2411387	Collaborative Research: Elements: VLCC-States: Versioned Lineage-Driven Checkpointing of Composable States	Rochester Institute of Technology
Ramanujan	Raghuram	2311263	Elements: Portable Machine Learning Models for Experimental Nuclear Physics	Davidson College
Randall	David	2005137	Collaborative Research: Frameworks: Community-Based Weather and Climate Simulation With a Global Storm-Resolving Model	Colorado State University
Rashidi	Abbas	2229603	Collaborative Research: CyberTraining: Implementation: Medium: CyberTraining of Construction (CyCon) Research Workforce Through an Educational and Community Engagement Platform	University of Utah
Roberts	Amy	2017760	CyberTraining: Implementation: Small: Enabling Dark Matter Discovery through Collaborative Cybertraining	CU Denver
Roberts	Amy	2104003	Collaborative Research: Elements: Shared Data-Delivery Infrastructure to Enable Discovery with Next Generation Dark Matter and Computational Astrophysics Experiments	CU Denver
Roberts	Amy	2411546	Collaborative Research: Elements: Shared Data-Delivery Infrastructure to Enable Discovery with Next Generation Dark Matter and Computational Astrophysics Experiments	CU Denver
Rodriguez	Meghan	2230127	CyberTraining: CIP: Training and Developing a Research Computing and Data CI Professionals (RCD-CIP) Community	Tufts University
Ruan	Xiulin	2311848	Elements: FourPhonon: A Computational Tool for Higher-Order Phonon Anharmonicity and Thermal Properties	Purdue University
Safronova	Marianna	2209639	Elements: Scalable and Automated Atomic Portal - Bridging the Gap Between Research Codes and User Community	University of Delaware
Samadi	Vidya	2320979	Collaborative Research: CyberTraining: Implementation: Small: Inclusive Cyberinfrastructure and Machine Learning Training to Advance Water Science Research	Clemson University
Saule	Erik	1924057	Collaborative Proposal: CyberTraining: Pilot: Aligning Learning Materials with Curriculum Standards to Integrate Parallel and Distributed Computing Topics in Early CS Education	UNC Charlotte
Schaffner	David	1931388	Collaborative Research: Frameworks: An open source software ecosystem for plasma physics	Bryn Mawr College
Schreiner	Henry	2209877	Elements: Simplifying Compiled Python Packaging in the Sciences	Princeton University
Schuchart	Joseph	1931387	Collaborative Research: Frameworks: Production quality Ecosystem for Programming and Executing eXtreme-scale Applications (EPEXA)	Stony Brook University
Shao	Yihan	2311442	Elements: An Integrated Software Platform for Simulating Polariton Photochemical and Photophysical Processes	University of Oklahoma
Shephard	Mark	2209471	Collaborative Research: Frameworks: A Software Ecosystem for Plasma Science and Space Weather Applications	Rensselaer Polytechnic Institute
Shephard	Mark	2209472	Collaborative Research: Frameworks: A Software Ecosystem for Plasma Science and Space Weather Applications	Rensselaer Polytechnic Institute
Shu	Tong	2306184	Collaborative Research: CyberTraining: Pilot: Research Workforce Development for Deep Learning Systems in Advanced GPU Cyberinfrastructure	University of North Texas
Simoneau	Eric	2118302	CyberTraining: Implementation: Medium: Collaborative Research: Computational and Data-Centric Ecology Training	33Sigma Learning Labs
Simoneau	Eric	2118305	CyberTraining: Implementation: Medium: Collaborative Research: Computational and Data-Centric Ecology Training	33Sigma Learning Labs
Simpson	Isla	2004575	Elements: Cyberinfrastructure for streamlining coupled simplified climate modeling within the Community Earth System Model	NSF NCAR
Simpson	Isla	2311376	Atmospheric Physics Needs for Community Climate Modeling	NSF NCAR
Sinkovits	Robert	2104104	Elements: Spatial Ecology Gateway	University of California, San Diego
Sinkovits	Robert	2320934	CyberTraining: Implementation: Small: COMPrehensive Learning for end-users to Effectively utilize CyberinfraStructure (COMPLECS)	University of California, San Diego
Sisneros	Robert	2209767	Collaborative Research: Elements: Towards A Scalable Infrastructure for Archival and Reproducible Scientific Visualizations	National Center for Supercomputing Applications
Snapp-Childs	Winona	2138307	Track 3: COre National Ecosystem for CyberinfrasTructure (CONECT)	Indiana University
Snapp-Childs	Winona	2227627	RCN:CIP: Midwest Research Computing and Data Consortium	Indiana University
Snapp-Childs	Winona	2321091	Collaborative Research: SCIPE: CyberInfrastructure Professionals InnoVating and brOadening the adoption of advanced Technologies (CI PIVOT)	Indiana University
Snapp-Childs	Winona	2321091	Collaborative Research: SCIPE: CyberInfrastructure Professionals InnoVating and brOadening the adoption of advanced Technologies (CI PIVOT)	Indiana University
Son	Seung Woo	2312982	OAC Core: Improving Data Integrity for HPC Datasets using Sparsity Profile	University of Massachusetts Lowell
Song	Carol	1835822	Framework: Data: HDR: Extensible Geospatial Data Framework towards FAIR (Findable, Accessible, Interoperable, Reusable) Science	Purdue University
Song	Yang	2230046	Collaborative Research: CyberTraining: Pilot: A Cybertraining Program to Advance Data Acquisition, Processing, and Machine Learning-based Modeling in Marine Science	UNC Wilmington
Song	Houbing	2309760	Collaborative Research: CyberTraining: Pilot: Operationalizing AI/Machine Learning for Cybersecurity Training	University of Maryland, Baltimore County
Speyer	Gil	2230108	A Cross-Institutional Research Engagement Network for CI Facilitators	Arizona State University
Srivastava	Ankit	2321005	CyberTraining: Implementation: Medium: Computational Materials Science Summer School - Fostering Accelerated Scientific Techniques (CMS3-FAST)	Texas AM University
Stamper	John	2209819	Collaborative Research: Frameworks for Intelligent Adaptive Experimentation: Enhancing and Tailoring Digital Education	Carnegie Mellon University
Stanzione	Dan	2321040	SCIPE: Chishiki.ai: A sustainable, diverse, and integrated CIP community for Artificial Intelligence in Civil and Environmental Engineering	The University of Texas at Austin
Stoeger	Thomas	2410335	Frameworks: Born Physical, Studied Digitally (BP/SD)	Northwestern University
Strachan	Scotty	2209806	CSSI: Elements: Innovating for Edge-to-Edge Climate Services	Nevada System of Higher Education
Street	Rachel	2209852	Frameworks: Target and Observation Manager Systems for Multi-Messenger and Time Domain Science	Las Cumbres Observatory
Sun	Ziheng	2117834	CyberTraining: Implementation: Medium: Machine Learning Training and Curriculum Development for Earth Science Studies	George Mason University
Sun	Wenhao	2209423	Elements: Data Driven Autonomous Thermodynamic and Kinetic Model Builder for Microstructural Simulations	University of Michigan
Sun	Hongyue	2230025	CyberTraining: Implementation: Small: Infrastructure Cybersecurity Curriculum Development and Training for Advanced Manufacturing Research Workforce	University of Georgia
Sussman	Alan	2017590	Collaborative Research:CyberTraining:Implementation:Medium: Broadening Adoption of Parallel and Distributed Computing in Undergraduate Computer Science and Engineering Curricula	University of Maryland
Sussman	Alan	2321015	Collaborative Research: CyberTraining:Implementation:Medium: Modern Course Exemplars infused with Parallel and Distributed Computing for the Introductory Computing Course Sequence	
Szewc	Manuel	2103889	Elements: Machine Learning Quark Hadronization	University of Cincinnati

PARTICIPANT LAST NAME	FIRST NAME	AWARD #	AWARD TITLE	Company
Taha	Ahmad	2230087	Collaborative Research: CyberTraining: Implementation: Medium: Cross-Disciplinary Training for Joint Cyber-Physical Systems and IoT Security	Vanderbilt University
Takabi	Daniel	2413654	CyberTraining: Implementation: Small: Building Future Research Workforce in Trustworthy Artificial Intelligence (AI)	Old Dominion University
Taylor	Greg	2103707	Collaborative Research: Elements: A CPU/GPU Pipeline Framework for High Throughput Data Acquisition and Analysis	The University of New Mexico
Thiruvathukal	George	2104319	CDSE: Collaborative: Cyber Infrastructure to Enable Computer Vision Applications at the Edge Using Automated Contextual Analysis	Loyola University Chicago
Thomas	Mary	2017767	CyberTraining: Developing a Best Practices Training Program in Cyberinfrastructure-Enabled Machine Learning Research	University of California at San Diego
Thomas	Mary	2230127	CIP: Training and Developing a Research Computing and Data CI Professionals (RCD-CIP) Community	University of California at San Diego
Tian	Yuan	2411153	Collaborative Research: Frameworks: MobilityNet: A Trustworthy CI Emulation Tool for Cross-Domain Mobility Data Generation and Sharing towards Multidisciplinary Innovations	University of California Los Angeles
Tovar	Benjamin	1931348	DataSwarm: User-Level Framework for Data Intensive Scientific Applications	University of Notre Dame
Turilli	Matteo	1931512	RADICAL-Cybertools: Middleware Building Blocks for NSF's Cyberinfrastructure Ecosystem	Rutgers University
Turilli	Matteo	2103986	Runtime for Heterogeneous Applications, Service Orchestration and DYnamism	Rutgers University
Van der Ven	Anton	2311370	Elements: Software to enable first-principles multi-scale simulations	University of California Santa Barbara
Verhagen	Marc	2104025	CSSI Elements: Towards a Robust Cyberinfrastructure for NLP-based Search and Discoverability over Scientific Literature	Brandeis University
Vieglais	Dave	2004815	Collaborative Research: Frameworks: Internet of Samples: Toward an Interdisciplinary Cyberinfrastructure for Material Samples	University of Kansas
Wade	Madeline	2103662	Framework: An A+ Framework for Multimessenger Astrophysics Discoveries through Real-Time Gravitational Wave Detection	Kenyon College
Walker-Loud	André	2311431	Collaborative Research: Elements: Lattice QCD software for nuclear physics on heterogeneous architectures	UC Berkeley / LBNL
Wang	Shaowen	2321070	Collaborative Research: CyberTraining: Implementation: Small: Broadening Adoption of Cyberinfrastructure and Research Workforce Development for Disaster Management	University of Illinois Urbana-Champaign
Wang	Shaowen	2411438	Collaborative Research: CSSI Frameworks: From Notebook to Workflow and Back Again	University of Illinois Urbana-Champaign
Wang	Chuang	2417717	Collaborative Research: CyberTraining: Implementation: Medium: AI4EDU: Cloud Infrastructure-Enabled Training for AI in Educational Research and Assessment	University of North Carolina at Charlotte
Wang	Jason	2320147, 2320148	Collaborative Research: CyberTraining: Pilot: Cyberinfrastructure-Enabled Machine Learning for Understanding and Forecasting Space Weather	New Jersey Institute of Technology
Wang	Guang	2411151, 2411152, 2411153	Collaborative Research: Frameworks: MobilityNet: A Trustworthy CI Emulation Tool for Cross-Domain Mobility Data Generation and Sharing towards Multidisciplinary Innovations	Florida State University
Wei	Tao	2118099	Collaborative Research: CyberTraining: Implementation: Medium: Cyber Training on Materials Genome Innovation for Computational Software (CyberMAGICS)	University of South Carolina
White	Joseph	2004954	Elements: Development and Dissemination of a Slurm Simulator	University at Buffalo
Wild	Stefan	2004601	Frameworks: Bayesian Analysis of Nuclear Dynamics	Northwestern University LBNL
Wilkinson	Michael	2103889	Elements: Machine Learning Quark Hadronization	University of Cincinnati
Wu	Yinghui	2104007	Elements: Crowdsourced Materials Data Engine for Unpublished XRD Results	Case Western Reserve University
Xu	Zhihan	2311870	Elements: Portable Library for Homomorphic Encrypted Machine Learning on FPGA Accelerated Cloud Cyberinfrastructure	University of Southern California
Xu	Dongkuan	2417850	Collaborative Research: CyberTraining: Implementation: Medium: EcoTern: Pioneering a CI Workforce for Sustainable and Transdisciplinary Environmental Science Research	North Carolina State University
Yang	Yuxin	2209563	OAC Core: Scalable Graph ML on Distributed Heterogeneous Systems	University of Southern California
Yang	Xianfeng	2234292	OAC Core: Stochastic Simulation Platform for Assessing Safety Performance of Autonomous Vehicles in Winter Seasons	University of Maryland: College Park
Yang	Lishan	2417718	Collaborative Research: CyberTraining: Implementation: Medium: AI4EDU: Cloud Infrastructure-Enabled Training for AI in Educational Research and Assessment	George Mason University
Yao	Yifan	2209857		University of Illinois, Urbana Champaign
Yu	Hui-Chia	2311466	Battery Electrode Simulation Toolkit using MFEM (BESFEM)	Michigan State University
Yu	Weikuan	2403089	Collaborative Research: OAC Core: CropDL - Scheduling and Checkpoint/Restart Support for Deep Learning Applications on HPC Clusters	Florida State University
Zabih	Ramin	2311521	Frameworks: arXiv as an accessible large-scale open research platform	Cornell Tech
Zender	Charlie	2004993	Elements: Advanced Lossless and Lossy Compression Algorithms for netCDF Datasets in Earth and Engineering Sciences (CANDEE)	University of California, Irvine
Zentner	Michael	2231406	CI CoE: SGX3 - A Center of Excellence to Extend Access, Expand the Community, and Exemplify Good Practices for CI Through Science Gateways	SDSC
Zhang	Liqing	2004751	Frameworks: Developing CyberInfrastructure for Waterborne Antibiotic Resistance Risk Surveillance (CI4-WARS)	Virginia Tech
Zhang	Liqing	2004751	Frameworks: Developing CyberInfrastructure for Waterborne Antibiotic Resistance Risk Surveillance (CI4-WARS)	Virginia Tech
Zhang	Xiaodong	2310510	Elements: Sustained innovation and service by a GPU-accelerated computation tool for applications of topological data analysis	The Ohio State University
Zhang	Xiangliang	2321054	C2D: Cybertraining for Chemical Data scientists	University of Notre Dame
Zhang	Zhe	2321069	Collaborative Research: CyberTraining: Implementation: Small: Broadening Adoption of Cyberinfrastructure and Research Workforce Development for Disaster Management	Texas AM University
Zhang	Zhao	2401245	Frameworks: Diamond: Democratizing Large Neural Network Model Training for Science	Rutgers University
Zhang	Zhao	2411294	Frameworks: hpcGPT: Enhancing Computing Center User Support with HPC-enriched Generative AI	Rutgers University
Zhao	Lan	2230092	Collaborative Research: CyberTraining: Implementation: Medium: Cyber Training for Open Science in Climate, Water and Environmental Sustainability	Purdue University
Zhao	Lan	2336630	Collaborative Research: CyberTraining: Implementation: Medium: Cyber Training for Open Science in Climate, Water and Environmental Sustainability	Purdue University
Zhao	Liang	2403312	Collaborative Research: OAC Core: Distributed Graph Learning Cyberinfrastructure for Large-scale Spatiotemporal Prediction	Emory University
Zhe	Shandian	2311685	Elements: A Convergent Physics-based and Data-driven Computing Platform for Building Modeling	University of Utah
Zhong	Zichun	2311245	Elements: MVP: Open-Source AI-Powered MicroVessel Processor for Next-Generation Vascular Imaging Data	Wayne State University
Zhou	Yang	2313191	Collaborative Research: OAC CORE: Federated Machine Learning for Rapid Traffic Event Detection and Response in Intelligent Transportation Systems	Auburn University
Zhuang	Yu	2321110	Collaborative Research: CyberTraining: Implementation: Small: Train the Trainers as Next Generation Leaders in Data Science for Cybersecurity for Underrepresented Communities	Computer Science Department, Texas Tech University
Bost	John	N/A	N/A	UNC-Chapel Hill
Cassarino	Nicholas	N/A	N/A	University of North Carolina at Charlotte
Cheng	Ming-Cheng	N/A	N/A	Clarkson University
Cohn-Gordon	Reuben	N/A	Elements: A new generation of samplers for astronomy and physics	Berkeley
Erpenbeck	Andre	N/A	N/A	University of Michigan
Gao	Jing	N/A	N/A	University of Delaware
Hakobyan	Hayk	N/A	N/A	Princeton
Heroux	Michael	N/A	N/A	Sandia National Labs
Hubbard	Rebecca	N/A	N/A	Tufts University
Jounaid	Soufiane	N/A	N/A	University of Chicago
Kolker-Hicks	Elliot	N/A	N/A	UNC Charlotte
Liu	Yuchen	N/A	N/A	North Carolina State University
Mahsa	Dabaghmeshin	N/A	N/A	University of Wisconsin-Milwaukee
Merola	Amanda	N/A	N/A	Tufts University
Milanese	Enrico	N/A	N/A	Woods Hole Oceanographic Institution
Panthi	Krishna	N/A	N/A	Clemson University
Rathbun	Susan	N/A	N/A	SDSC
Wang	Haoxin	N/A	N/A	Georgia State University
Wang	Jinzhen	N/A	N/A	UNC Charlotte
Wang	Yang	N/A	N/A	Carnegie Mellon University

AUTHORS: A to Z	AWARD#	POSTER TITLE	figshare DOI
Name	12345	TItle	figshareDOI
Agar, Joshua	2209135	Cyberinfrastructure for Scientific Data Preservation and Image Similarity Search	10.6084/m9.figshare.26798377
Akimov, Alexey	1924256	CyberTraining: Pilot: Modeling Excited State Dynamics in Solar Energy Materials	10.6084/m9.figshare.26532907.v1
Akimov, Alexey	1931366	Elements: Libra: The Modular Software for Nonadiabatic and Quantum Dynamics	10.6084/m9.figshare.26532907.v1
Akli, Linda	2400201	Computational and Data Science Curriculum Exchange (CDS-Exchange)	10.6084/m9.figshare.26517091.v1
Aksamija, Zlatan	2302879	Thermal and Electronic Transport Dynamics in 2D Materials	10.6084/m9.figshare.26800810.v1
Amrhein, Dan	2311382	CROCODILE: Facilitating rapid prototyping and analysis of regional ocean physics and carbon cycle modeling with data assimilation	10.6084/m9.figshare.26820904
Andreussi, Oliviero	2321102	The Quantum-MultiScale Collaboration for CyberTraining	10.6084/m9.figshare.26543785
Arora, Ritu	2314203	Basil: A Tool for Semi-Automatic Containerization, Deployment, and Execution of Scientific Applications on Cloud Computing and Supercomputing Platforms	10.6084/m9.figshare.26488540.v1
Arratia, Miguel	2311667	Infrastructure for Unbinned Unfolding (Deconvolution)	10.6084/m9.figshare.26801119
Atherton, Timothy	2003820	Morpho—A programmable environment for shape optimization and shapeshifting problems	10.6084/m9.figshare.28143788
Aydin, Berkay	2104004	Enhancing Space Weather Forecasting Capabilities with Shape-based Active Region Characteristics	10.6084/m9.figshare.26543326
Barton, Michael	2103905	Integrative Cyberinfrastructure for Next-Generation Modeling Science	10.6084/m9.figshare.26524618
Bass, Steffen	2004571	Probing excited nuclear matter with a Statistically and Computationally Advanced Program Envelope (X-SCAPE)	10.6084/m9.figshare.26511046
Bhasi, Vivek	1931531	Empowering Galaxy via Heterogeneous Compute and Storage	10.6084/m9.figshare.26548294
Bhatia, Sajal	2017371	Problem-Based Learning Approach for Ethical Hacking and Network Defense	10.6084/m9.figshare.26536270
Bhowmick, Sanjukta	2104076	Analyzing Dynamic Networks with CANDY	10.6084/m9.figshare.26800732
Bockelman, Brian	2209645	Enabling Storage Management Policies on the OSDF	10.6084/m9.figshare.26599945
Brinson, Cate	1835677	Nanocomposites to Metamaterials: A Knowledge Graph Framework	10.6084/m9.figshare.26799556
Candan, K. Selcuk	2311716	CausalBench: A Cyberinfrastructure for Causal-Learning Benchmarking for Efficacy, Reproducibility, and Scientific Collaboration	10.6084/m9.figshare.26662369
Carifio, Jonathan	2209623, 2209624, 2209625	Collaborative Research: Elements: Enriching Scholarly Communication with Augmented Reality	10.6084/m9.figshare.26587411
Carr, Steve	2320951	Promoting AI Readiness for Machine-Assisted Secure Data Analysis	10.6084/m9.figshare.26490604.v1
Carver, Jeff	2017259, 2017424	INnovative Training Enabled by a Research Software Engineering Community of Trainers (INTERSECT)	10.6084/m9.figshare.26446243
Casanova, Henri	2103489	Simulation-driven Evaluation of Cyberinfrastructure Systems	10.6084/m9.figshare.26419927
Chakraborty, Ankit	2103524	Elements:Software A Scalable Open-Source hp-Adaptive FE Software for Complex Multiphysics Applications	10.6084/m9.figshare.26535712
Chakravorty, Dhruva	2320992	ByteBoost: Building a strong community of computational researchers empowered in the use of novel cutting-edge technologies	MISSING
Chard, Kyle	2004894 , 2004932	Globus Compute: Federated Function as a Service for Research Cyberinfrastructure	10.6084/m9.figshare.26798602
Chard, Kyle	2104013	ChronoLog: A distributed shared log storage	10.6084/m9.figshare.26798602
Chaudhary, Vipin	2320952	Interdisciplinary Research Support Community for Artificial Intelligence and Data Sciences	MISSING
Cheng, Ming-Cheng	2003307	Rigorous Projection-based Learning Guided by Physical Principles for Fast and Accurate Dynamic Thermal Simulations of Multicore Microprocessors	10.6084/m9.figshare.26621809.v1
Chilton, Lydia	2103794	DESDR: Open-Source Cyberinfrastructure as a Decision Engine for Socioeconomic Disaster Risk	10.6084/m9.figshare.26517583.v1
Choi, Eunseo	2104002	Latest Progress in Developing DES3D, an Integrated Modeling Platform for Tectonics	10.6084/m9.figshare.26520877.v1
Chourasia, Amit	2311206, 2311207, 2311208	Collaborative Research: Frameworks: Quakeworx - An extensible software framework for earthquake simulations	10.6084/m9.figshare.26741017.v1
Chourasia, Amit	2231406	SGX3 - The Science Gateways Center of Excellence: A Professional Partner for Impactful Cyberinfrastructure	10.6084/m9.figshare.26741017.v1
Cleveland, Sean	2118222	Cyberinfrastructure training to Advance Environmental Science CI-TRACS Implementation: Medium Program	10.6084/m9.figshare.26810521
Cleveland, Sean	1931575	Collaborative Framework: Project Tapis: Next Generation Software for Distributed Research	10.6084/m9.figshare.26810521
Closser, Kristina	2417553	ChemCompute: Computational Chemistry for Undergraduate Education and Research	10.6084/m9.figshare.26767942.v1
Codoni, David	2310909	A multi-fidelity computational framework for vascular mechanobiology in SimVascular	10.6084/m9.figshare.26548309
Cohn-Gordon, Reuben	2311559	A new generation of Monte Carlo samplers for astronomy and physics	10.6084/m9.figshare.26799196
Coles, Victoria	2321008	CGC-SCIPE: Enhancing the Transdisciplinary Research Ecosystem for Earth and Environmental Science with Dedicated Cyber Infrastructure Professionals	10.6084/m9.figshare.26767684
Corbin, Nicole	2319979	Machine Learning Foundations and Applications in the Earth Systems Sciences	10.6084/m9.figshare.26799808
Crawford, Daniel	2136142	The Molecular Sciences Software Institute	10.6084/m9.figshare.26798539
Crawford, Daniel	2321044	Accelerating Curricular Transformation in the Computational Molecular Sciences	10.6084/m9.figshare.26798539
Crichigno, Jorge	2118311	Cybertraining on P4 Programmable Devices using an Online Scalable Platform with Physical and Virtual Switches and Real Protocol Stacks	10.6084/m9.figshare.26798215
Crichigno, Jorge	2403360	Enhancing Network Security by Implementing an ML Malware Detection and Classification Scheme in P4 Programmable Data Planes and SmartNICs	10.6084/m9.figshare.26798215
Crosby, Lonnie	2230106; 2230108	A Cross-Institutional Research Engagement Network for CI Facilitators	10.6084/m9.figshare.26756998
Crozier, Peter	2104105	Unsupervised Denoising for Atomic Resolution Imaging and Spectroscopy	10.6084/m9.figshare.26801143
Cui, Hantao	2319895	PowerCyber Training: Computational Training for Power Engineering Researchers	MISSING

AUTHORS: D to K	AWARD#	POSTER TITLE	figshare DOI
D'Orazio, Vito	2311142	Frameworks: Infrastructure For Political And Social Event Data using Machine Learning	10.6084/m9.figshare.26720437
Dabaghmeshin, Mahsa	2229652	CIberCATSS: A Comprehensive, Applied and Tangible CyberInfrastructure Summer School in Southeastern Wisconsin	10.6084/m9.figshare.26513080
Davidson, Drew	2311819	Prototyping a Quality Control / Quality Assurance System for the StraboSpot Geologic Information System	10.6084/m9.figshare.26767156.v1
DePrince, Eugene	2103705	Unitary Coupled Cluster in the Chronus Quantum Package	10.6084/m9.figshare.26768125.v1
Denolle, Marine	2103701	SCOPED Update: a Cloud and HPC software platform for computational seismology	10.6084/m9.figshare.26513095.v2
Deppe, Nils	2209655	SpECTRE: A task-based code for multiphysics problems in astrophysics at exascale	10.6084/m9.figshare.26798689
Di, Sheng	2311875	FZ: A fine-tunable cyberinfrastructure framework to streamline specialized lossy compression development	10.6084/m9.figshare.26359654.v1
Di, Sheng	2104023,2247080	Collaborative Research: Elements: ROCCI: Integrated Cyberinfrastructure for In Situ Lossy Compression Optimization Based on Diverse User Requirements	MISSING
Dong, Wenqian	2417849	MISSING	MISSING
Dunbar, Oliver	1835860	Data-driven ocean, atmosphere, and land parameterizations calibrated from indirect data	10.6084/m9.figshare.26540629.v2
Eisma, Jessica	2230054	Justice in Data: Developing FAIR-aligned data science skills for undergraduate water and energy researchers	10.6084/m9.figshare.26799190
Esfarjani, Keivan	2103989	ALADYN: suites of code based on Anharmonic LAttice DYNamics to model thermal properties of materials	10.6084/m9.figshare.26540689.v1
Evans, John	2104106	Collaborative Research: Elements: EXHUME: Extraction for High-Order Unfitted Finite Element Methods	10.6084/m9.figshare.26585050
Figueiredo, Renato	2311123	FaaSr: R package for Function-as-a-Service Cloud Computing	10.6084/m9.figshare.26381704
Ganapati, Sukumar	1924154	Cyberinfrastructure Training for Policy Scientists	10.6084/m9.figshare.26827435
Gates, Mark	2004541	Advancing Linear Algebra Capabilities in the BALLISTIC Project	10.6084/m9.figshare.26759863
Gerber, Edwin	2004572	Towards a Data-driven Representation of Un(der)resolved Gravity Waves in Atmospheric Models	10.6084/m9.figshare.26543812
Giustino, Feliciano	2103991	Many-body electronic structure calculations on the cloud using MATCSSI	10.6084/m9.figshare.26513296
Griffith, Boyce	1931516	Multiphase Fluid-Structure Interaction Software Infrastructure to Enable Applications in Medicine, Biology, and Engineering	10.6084/m9.figshare.26828260.v1
Guan, Qiang	2230111	Interactive and Integrated Training for Quantum Application Developers across Platforms	10.6084/m9.figshare.28138184
Gull, Emanuel	2310582	The GREEN Package: Embedding Framework for Quantum Many-Body Simulations	10.6084/m9.figshare.26798536
Guo, Yuebin	CMMI2152908	Physics-Informed Deep Learning of Melt Pool Dynamics	MISSING
Haas, Roland	2004879	The Einstein Toolkit ecosystem: Enabling fundamental research in the era of multi-messenger astrophysics	10.6084/m9.figshare.26543776
Haas, Roland	2310548	Elements: An initial value solver for the era of multi-messenger astrophysics	10.6084/m9.figshare.26764174
Hamilton, Kathryn	2311928	The AMOS Gateway: An Advanced Cyberinfrastructure for Atomic, Molecular, and Optical Science	10.6084/m9.figshare.26523961.v1
He, Suining	2118102	CyberTraining: Pilot: Cyberinfrastructure Training in Computer Science and Geoscience	10.6084/m9.figshare.26535730
Heffernan, Neil	1931523	Collaborative Research: Frameworks: Cyber Infrastructure for Shared Algorithmic and Experimental Research in Online Learning - Update	MISSING
Heinz, Hendrik	1931587	CSSI Framework: Cyberloop for Accelerated Bionanomaterials Design	10.6084/m9.figshare.26541124
Heyden, Matthias	2311372	Analyzing molecular simulations on-the-fly using a streaming interface	10.6084/m9.figshare.26662492.v1
Hillery, Elizabett (Betsy)	2321090	CI-PIVOT: A Novel Approach to Expanding the CI Professional Ecosystem	10.6084/m9.figshare.26809261.v
Howard, Michael	2310725	Multiparticle collision dynamics simulations of mesoscale hydrodynamic interactions in complex soft materials and environments	10.6084/m9.figshare.26662795
Hu, Ming	2311202	Phonon Database Generation, Analysis, and Visualization for Data Driven Materials Discovery	10.6084/m9.figshare.26740477
Hudak, David	1835725	Open OnDemand: Connecting Computing Power With Powerful Minds	10.6084/m9.figshare.26524600.v1
Hutton, Eric	2104102	OpenEarthscape: Tools and resources for community modeling of earth-surface dynamics	10.6084/m9.figshare.26801029.v1
Jafarzadeh, Seyed	2311527	Widefield Images Stellar Photometry	10.6084/m9.figshare.26524369.v1
Jawed, Khalid	2209782	DisMech: Discrete Simulation of Flexible Structures and Soft Robots	10.6084/m9.figshare.26799004
Jee, Kangkook	2321117	Expanding Research in Space System Safety and Security: Educational Initiatives and Future Directions	MISSING
Jiang, Weiwen	2311949; 2320957	Quantum Pathway to Quantum Education and Quantum Utility	10.6084/m9.figshare.26798884.v2
Jounaid, Soufiane	2230077	FOUNT: Connecting Code and Hardware in CI Education	MISSING
Kalyanam, Rajesh	2311762	Science-i Cyberinfrastructure for Forest Ecosystem Research	10.6084/m9.figshare.26338960
Kanai, Yosuke	2209857/2209858	Tackling the Accuracy and Efficiency Bottlenecks in Real-Time TDDFT Simulation	MISSING
Kandasamy, Nagarajan	2209745	A Fully Configurable Open-Source Software-Defined Digital Quantized Spiking Neural Core Architecture	10.6084/m9.figshare.26577538
Karagiorgi, Georgia	2209917	Elements: RAD Discoveries for Fundamental Physics	10.6084/m9.figshare.26781502
Katsavounidis, Erik	1931469	ML workflows for real-time detection and parameter estimation of gravitational-wave sources	10.6084/m9.figshare.26540761.v1
Katz, Daniel	2209920	Sustaining Parsl: Productive Parallel Programming in Python	10.6084/m9.figshare.26661619.v1
Kerzendorf, Wolfgang	2311323	The TARDIS radiative transfer framework	10.6084/m9.figshare.26831749.v1
Kim, Hyesoon	2103951	Open-source hardware and software evaluation system for UAV	10.6084/m9.figshare.26800834
Kline Struhl, Melissa	2209756	Mobile data collection and open source infrastructure for remote data collection with babies and children	10.6084/m9.figshare.26530885.v2
Knepper, Rich	2320977	HPC-ED Federated Learning Catalog for Scientific Computing Training Materials	10.6084/m9.figshare.26513359
Knepper, Rich	2209711	CSSI: Frameworks: Large Scale Atmospheric Research Using an Integrated WRF Modeling, Visualization, and Verification Framework (I-WRF)	10.6084/m9.figshare.26513359
Knight, Christopher	2311260	Frameworks: Data-Driven Software Infrastructure for Next-Generation Molecular Simulations	10.6084/m9.figshare.26782720.v1
Kumar, Krishna	2103937	Cognitasium - Enabling Data-Driven Discoveries in Natural Hazards Engineering	MISSING
Kumar, Ratnesh	2004766	En-MISSION: Environment-aware Model-based In-Season Sensor-driven Scheduling of Irrigation Or/and Nitrogen	10.6084/m9.figshare.26814892
Kumar, Krishna	2321040	AI developments in Civil Engineering	MISSING

AUTHORS: L to O	AWARD#	POSTER TITLE	figshare DOI
Lange, David	2311471	Clad: Enabling Differentiable Programming in Science	10.6084/m9.figshare.26780410
Lejeune, Emma	2310771	Curating and Disseminating Solid Mechanics Based Benchmark Datasets	10.6084/m9.figshare.26540602
Lele, Sanjiva	2103509	Compack3D: Accelerating high-order compact scheme simulations	10.6084/m9.figshare.28138154
Lemson, Gerard	2311791	Open SciServer: A Sustainable Data-Driven Science Platform	10.6084/m9.figshare.26568826.v1
Lewis, Thu-Mai	2209628, 2209629, 2209630	TRAnsparency CErtified (TRACE): Trusting Computational Research Without Repeating It	10.6084/m9.figshare.26530711
Li, Xiaosong	2103717	Sustainable Open-Source Quantum Dynamics and Spectroscopy Software	10.6084/m9.figshare.26768125.v1
Li, Dong	A210125002	Accelerating LLM Inference Using Memoization in SciMem	10.6084/m9.figshare.26384722
Liang, Xu	2209833, 2209835, 2209834	Collaborative Research: Frameworks: Building a collaboration infrastructure: CyberWater2—A sustainable data/model integration framework	10.6084/m9.figshare.26799199
Liang, Xin	2313122/2313123/2313124	OAC Core: Topology-Aware Data Compression for Scientific Analysis and Visualization	10.6084/m9.figshare.26618158
Liang, Xin	2311756/2311757/2311758	ProDM: Developing A Unified Progressive Data Management Library for Exascale Computational Science	10.6084/m9.figshare.26618158
Lindner, Peggy	2321110	Curriculum to Train the Trainers as Next Generation Leaders in Data Science for Cybersecurity	10.6084/m9.figshare.26798431
Lindzey, Laura	2209726	QIceRadar: A QGIS plugin for finding, downloading, and visualizing georeferenced radargrams	10.6084/m9.figshare.26796136
Liu, Dunyu	2209807	Monitoring Earth Surface Deformation with the Next Generation of InSAR Satellites: GMTSAR	10.6084/m9.figshare.26767699.v1
Liu, Qianqian	2230046	A Cybertraining Program to Advance Data Acquisition, Processing, and Machine Learning-based Modeling in Marine Science	10.6084/m9.figshare.26496496
Logan, Luke	2313154	LABIOS: Storage Acceleration via Data Labeling and Asynchronous I/O	10.6084/m9.figshare.26809108
Lou, Helen	2321055	Multidisciplinary Cybersecurity Education for Critical Energy and Chemical Infrastructure Resilience	10.6084/m9.figshare.26800933.v1
Lowe-Power, Jason	2311888	Advancing Computer Hardware and Systems' Research Capability, Reproducibility, and Sustainability with the gem5 Simulator Ecosystem	10.6084/m9.figshare.26540617.v1
Lu, Guoyu	2334690	3D Modeling and Assessment Infrastructure For Agriculture and Robotics Applications	10.6084/m9.figshare.26524783
Lu, Xiaoyi	2321123	CyberTraining: Pilot: Cross-Layer Training of High-Performance Deep Learning Technologies and Applications for Research Workforce Development in Central Valley	10.6084/m9.figshare.26801380
Lv, Qin	2311843	Sonar AI: Development of cyberinfrastructure to establish a scalable application of self-supervised machine learning for over a decade of NOAA's water column sonar data	10.6084/m9.figshare.26662705
Lynch, Benjamin	2320769	Cyber Training: Pilot – Breaking the Compute Barrier, Upskilling Agri-Food Researchers to Utilize HPC Resources	10.6084/m9.figshare.26510716.v1
Machi, Dustin	1916805	CINES: A Scalable Cyberinfrastructure for Sustained Innova- tion in Network Engineering and Science	10.6084/m9.figshare.26808634
Maffeo, Christopher	2311550	Multi-resolution modeling of biology and nanotechnology systems with ARBD	10.6084/m9.figshare.26662693
Mandal, Subhasish	2311557,2311558	Towards building an open-source DFT+eDMFT database for quantum materials	10.6084/m9.figshare.26630545
Mandal, Anirban	2320373	CyberInfrastructure Training and Education for Synchrotron X-Ray Science (X-CITE)	10.6084/m9.figshare.26487595
Manning, T. Andrew	2103680	Frameworks: MUSES, Modular Unified Solver of the Equation of State	10.6084/m9.figshare.26543767
Marru, Suresh	2209872	Cybershuttle: Advancing Science through a Computing Continuum	10.6084/m9.figshare.26798275
Martiniani, Stefano	2311632	An Integrated Platform for Training and Deploying Machine Learning Interatomic Potentials	10.6084/m9.figshare.26798545
May, Ryan	2103682	Scaling MetPy to Big Data Workflows in Meteorology and Climate Science	10.6084/m9.figshare.26801254
McCully, Curtis	2311355	SCIMMA: Real-time Orchestration of Multi-Messenger Astrophysical Observations	10.6084/m9.figshare.26672584
McHenry, Kenton	1835834	Open Source Data Management for Long Tail Data	10.6084/m9.figshare.26866708
McHenry, Kenton	2209863	DeCODER: Democratized Cyberinfrastructure for Open Discovery to Enable Research	10.6084/m9.figshare.26866708
Meneveau, Charles	210387	Advanced Cyberinfrastructure for Sustainable Community Usage of Big Data from Numerical Fluid Dynamics Simulations	10.6084/m9.figshare.26543641.v1
Menzies, Tim	1931425	When Will You Software Fail?	MISSING
Merz, Kenneth	2209717	CSSI: Frameworks: Interoperable High-Performance Classical, Machine Learning and Quantum Free Energy Methods in AMBER	10.6084/m9.figshare.26488039.v1
Mills, Katie	2311919	Argovis: A Data Service for Rapidly Searching and Cross-Referencing Ocean Data	MISSING
Mirkouei, Amin	2229604	CyberTraining of Construction (CyCon) Research Workforce Through an Educational and Community Engagement Platform	10.6084/m9.figshare.26540875.v1
Moore, Shirley	2311707	SPADE: Scalable Performance and Accuracy analysis for Distributed and Extreme-scale systems	10.6084/m9.figshare.26452465.v1
Moreno-Centeno, Erick	1835499	Sparse Exact (SPEX) LU and Cholesky Factorization Library	10.6084/m9.figshare.26800132
Morgan, Dane	2017072	Authentic, Accessible, Personalized, and Scalable Undergraduate Research in Data Science	10.6084/m9.figshare.26611246
Morgan, Dane	1931298	Cloud Hosting and Uncertainty Quantification for Machine Learning Data and Models in Materials	10.6084/m9.figshare.26611246
Morse, David	2103627	Elements: Open-source tools for block polymer phase behavior	10.6084/m9.figshare.26826529
Nachman, Benjamin	2311666	Enabling Particle and Nuclear Physics Discoveries with Neural Deconvolution	MISSING
Nakano, Aiichiro	2118061	CyberMAGICS: Cyber Training on Materials Genome Innovation for Computational Software	10.6084/m9.figshare.26512612
Narayanan, Sri Hari Krishna	2104068	DJ4Earth: Differentiable programming in Julia for Earth system models	10.6084/m9.figshare.26577739
Nawab, Faisal	2321121	Cybertraining to Develop FAIR Data Competencies for Bioengineering Students	10.6084/m9.figshare.26827228
Neeman, Henry	2118193	The Certified Cyberinfrastructure Facilitator Training and Development (CCIFTD) Program	10.6084/m9.figshare.26524612
Negrut, Dan	2209791	Simulating Autonomous Agents and the Human-Autonomous Agent Interaction	10.6084/m9.figshare.26824288
Ni, Zhen	1949921	Collaborative Research: CyberTraining: Implementation: Small: Multi-disciplinary Training of Learning, Optimization and Communications for Next Generation Power Engineers	10.6084/m9.figshare.26815645
Ni, Zhen	1949921	Collaborative Research: CyberTraining: Implementation: Small: Multi-disciplinary Training of Learning, Optimization and Communications for Next Generation Power Engineers	10.6084/m9.figshare.26815648
Nicolae, Bogdan	2411386	VLCC-States: Versioned Lineage-Driven Checkpointing of Composable States	MISSING
Niu, Wei	2403088, 2403089, 2403090	Collaborative Research: OAC Core: CropDL - Scheduling and Checkpoint/Restart Support for Deep Learning Applications on HPC	10.6084/m9.figshare.26527438
North, Chris	2004014,2003800,2003387	SAGE3 : Space to Think for Collaborative Research and Education	10.6084/m9.figshare.26360038
Olaya, Paula	2103845	End-to-end Integration of Fine-grained Environmental Workflows on HPC and Cloud Converged Infrastructure	10.6084/m9.figshare.26543791.v1
Ou, Ge	2004658	Elements: Open Access Data Generation Engine for Bulk Power System under Extreme Windstorms	MISSING

AUTHORS: P to S	AWARD#	POSTER TITLE	figshare DOI
Pallickara, Shrideep	1931363, 1931324, 1931335,,1931283	SUSTAIN: Catalyzing Urban Sustainability At Scale	MISSING
Panda, Dhabaleswar	2311830	CSSI Frameworks: Performance engineering scientific applications with MVAPICH and TAU using emerging communication primitives	10.6084/m9.figshare.26841616
Panda, Dhabaleswar	2007991	OAC Core: Next-generation communication and I/O middleware for HPC and deep learning with smart NICs	10.6084/m9.figshare.26841616
Pantano-Rubino, Carlos	2320943	Fostering Computational Excellence (FOCEX): Addressing the Disconnect between Advanced CyberInfrastructure and Educational Preparedness	10.6084/m9.figshare.26543671
Park, Seung-Jong	2403399	Harvesting Idle Resources Safely and Timely for Large-scale AI Applications in High Performance Computing Systems	10.6084/m9.figshare.26499400.v1
Patel, Vivak	2309445	A cyberlaboratory for randomized linear algebra	10.6084/m9.figshare.26591698
Patra, Abani	2004826	Ghub : Building a glaciology gateway to unify a community	MISSING
Peng, Shiyu	2209262	Elements: The PERTURBO Package: A Community Code for Electron Interactions and Dynamics in Materials	10.6084/m9.figshare.26688241
Philippov, Alexander	2311800	Entity: next-generation radiative particle-in-cell code for relativistic plasma astrophysics	10.6084/m9.figshare.26533300.v1
Phillips, Daniel	2004601	The Bayesian Analysis of Nuclear Dynamics CI Framework	10.6084/m9.figshare.26760496
Piper, Mark	2118171	Code to Communicate (CoCo)	10.6084/m9.figshare.26739088
Pivarski, Jim	2103945	Awkward Arrays - Accelerating scientific data analysis on irregularly shaped data	10.6084/m9.figshare.26543308.v1
Prasad, Sushil	2017590	What should every Computer Science and Computer Engineering Student know about Parallel and Distributed Computing?	10.6084/m9.figshare.26513230
Prasad, Sushil	2321015	CyberTraining: Modern Course Exemplars infused with Parallel and Distributed Computing for the Introductory Computing Course Sequence (Implementation, Medium)	10.6084/m9.figshare.26513230
Price-Skelly, Owen	2209892	Frameworks: Garden: A FAIR Framework for Publishing and Applying AI Models for Translational Research in Science, Engineering, Education, and Industry	10.6084/m9.figshare.26543686
Puri, Satish	2402987	Efficient Indexing and Similarity Search Systems on Heterogeneous Platforms	10.6084/m9.figshare.26530735
Purwanto, Wirawan	2320998	T3-CIDERS: Fostering a Community of Practice in CI- and Data-Enabled Cybersecurity Research Through A Train-the-Trainer Program	10.6084/m9.figshare.26589754
Quinn, Daven	2311091	Establishing a community-based data workflow to integrate stratigraphic and geochemical descriptions of the Earth's crust	10.6084/m9.figshare.26730637
Rai, Neeraj	2118204	Establishing Sustainable Ecosystem for Computational Molecular Science Training and Education	10.6084/m9.figshare.26543773
Rajib, Adnan	2230092,2336630	Cyber Training for Open Science in Climate, Water and Environmental Sustainability	MISSING
Ramanujan, Raghuram	2311263	Pre-trained Machine Learning Models to Support Nuclear Science	10.6084/m9.figshare.26767795.v1
Randall, David	2005137	A Global Storm Resolving Configuration of the Community Earth System Model	10.6084/m9.figshare.26799463
Rashidi, Abbas	2229603	CyberTraining of Construction (CyCon) Research Workforce Through an Educational and Community Engagement Platform	10.6084/m9.figshare.26506654.v1
Ruan, Xiulin	2311848	FourPhonon: A Computational Tool for Higher-Order Phonon Anharmonicity and Thermal Properties	MISSING
Safronova, Marianna	2209639	Scalable and Automated Atomic Portal - Bridging the Gap Between Research Codes and User Community	10.6084/m9.figshare.26799031
Saidi , Wissam	2003808	Accelerating Atomistic Simulations are Extreme Conditions	MISSING
Samadi, Vidya	2320979	Best Practices in Water Science Data Analytics: Lesson learned from CyberTraining Project	10.6084/m9.figshare.26801080
Saule, Erik	1924057	CS Materials: Understanding your Classes and Improving CS Education	10.6084/m9.figshare.26510827
Schaffner, David	1931388	Collaborative Research: Frameworks: An open source software ecosystem for plasma physics	10.6084/m9.figshare.26799373
Schreiner, Henry	2209877	Elements: Simplifying Compiled Python Packaging in the Sciences	10.6084/m9.figshare.26536720
Schuchart, Joseph	1931387	CSSI:EPEXA: Ecosystem for Programming and Executing eXtreme-scale Applications	10.6084/m9.figshare.26540623.v1
Shao, Yihan	2311442	Elements: An Integrated Software Platform for Simulating Polariton Photochemical and Photophysical Processes	10.6084/m9.figshare.26825395.v1
Shephard, Mark	2209471	Cyberinfrastructure for Plasma Science and Space Weather Simulations	10.6084/m9.figshare.26364754.v2
Shu, Tong	2306184	Research Workforce Development for Deep Learning Systems in Advanced GPU Cyberinfrastructure	10.6084/m9.figshare.26768296
Simoneau, Eric	2118302	Data4Ecology.org — Facilitating Computational and Quantitative Skills in Undergraduate Ecology Courses	10.6084/m9.figshare.28152800
Simpson, Isla	2311376	Towards incorporating legacy versions of the Community Atmosphere Model within the Common Community Physics Package (CCPP) framwork.	10.6084/m9.figshare.26767498
Simpson, Isla	2004575	A graphical user interface for configuring idealized experimental setups with the Community Earth System Model	10.6084/m9.figshare.26767498
Sinkovits, Robert	2104104	Space Use Ecology Gateway	10.6084/m9.figshare.26608780.v1
Sinkovits, Robert	2320934	COMPLECS: COMPrehensive Learning for end-users to Effectively utilize CyberinfraStructure	10.6084/m9.figshare.26608780.v1
Sisneros, Robert	2209768	Elements: Towards A Scalable Infrastructure for Archival and Reproducible Scientific Visualizations	10.6084/m9.figshare.26767213
Snapp-Childs, Winona	2227627	Midwest Research Computing and Data Consortium	10.6084/m9.figshare.26798239
Son, Seung Woo	2312982	Estimating Silent Data Corruption Rates Using Hardware Counters	10.6084/m9.figshare.26662687
Song, Carol	1835822	GeoEDF: an Extensible Geospatial Data Framework towards FAIR	10.6084/m9.figshare.26761744.v1
Song, Houbing	2309760, 2229976	CyberTraining: Pilot: Operationalizing AI/Machine Learning for Cybersecurity Training	10.6084/m9.figshare.26517145
Srivastava, Ankit	2321005	Computational Materials Science Summer School - Fostering Accelerated Scientific Techniques	10.6084/m9.figshare.26746330
Stamper, John	2209819	Making Adaptive Experimentation EASI	10.6084/m9.figshare.26798260.v1
Stoeger, Thomas	2410335	A digital archive of a funding agency reveals how it cooperated with academics to support scientific innovation in a nascent field of science.	10.6084/m9.figshare.26768224.v1
Strachan, Scotty	2209806	Research and Development for the NevWx Edge-to-Edge Climate Services System	10.6084/m9.figshare.28143788
Street, Rachel	2209852	TOM Toolkit: Enabling follow-up of Time Domain and Multi-Messenger Astrophysical Discoveries	10.6084/m9.figshare.26539810
Sun, Wenhao	2209423	AMMBER: The AI-enabled Microstructure Model BuildER	10.6084/m9.figshare.26876704
Sun, Ziheng	2117834	GeoSMART: Open Curriculum for Bootstrapping AI Adoption in Earth Sciences	MISSING
Sun, Hongyue	2230025	CyberTraining: Implementation: Small: Infrastructure Cybersecurity Curriculum Development and Training for Advanced Manufacturing Research Workforce	10.6084/m9.figshare.26406319
Szewc, Manuel	2103889	Modeling Hadronization with Machine Learning	10.6084/m9.figshare.26539834

AUTHORS: T to Z	AWARD#	POSTER TITLE	figshare DOI
Taha, Ahmad	2230087	Stealthy False Data Injection Attacks on Drinking Water and Power Systems	10.6084/m9.figshare.26662243
Tajkhorshid, Emad	2139536	Boosting the Performance of NAMD on GPU-enabled Platforms	MISSING
Takabi, Daniel	2413654	Building Future Research Workforce in Trustworthy Artificial Intelligence (AI)	10.6084/m9.figshare.26799499
Tang, Yufei	2320972	Secure, Resilient Cyber-Physical Energy System Workforce Pathways via Data-Centric, Hardware-in-the-Loop Training	10.6084/m9.figshare.28138145
Taylor, Greg	2103707	Using Bifrost for High Speed GPU Imaging Pipelines	10.6084/m9.figshare.26536159
Thiruvathukal, George	2104319	Cyber Infrastructure to Enable Computer Vision Applications at the Edge Using Automated Contextual Analysis	10.6084/m9.figshare.26530858
Thiruvathukal, George	2107020	Advancing Low-Power Computer Vision at the Edge	10.6084/m9.figshare.26530858
Thomas, Mary	2017767	Best Practices Training Program in Cyberinfrastructure-Enabled Machine Learning (CIML)	10.6084/m9.figshare.29474351
Thomas, Mary	2230127, 2017767	CIP-Fellows: Training and Developing Research Computing and Data CI Professionals	10.6084/m9.figshare.29474351
Tovar, Benjamin	1931348	TaskVine: A User-Level Framework for Data Intensive Scientific Applications	10.6084/m9.figshare.26540632
Turilli, Matteo	1931512	RADICAL-Cybertools: Middleware Building Blocks for NSF's Cyberinfrastructure Ecosystem	10.6084/m9.figshare.26540734
Turilli, Matteo	2103986	RHAPSODY: Runtime for Heterogeneous Applications, Service Orchestration and DYnamism	10.6084/m9.figshare.26540734
Van der Ven, Anton	2311370	Surrogate models to enable multi-scale modeling of alloys for Li-ion batteries	10.6084/m9.figshare.26767807
Verhagen, Marc	2104025	CSSI Elements: Towards a Robust Cyberinfrastructure for NLP-based Search and Discoverability over Scientific Literature	10.6084/m9.figshare.26758540
Vieglais, Dave	2004815	iSamples (Internet of Samples): Cyberinfrastructure to support transdisciplinary use of material samples	10.6084/m9.figshare.26528062.v1
Wade, Madeline	2103662	An A+ framework for multimessenger astrophysics discoveries through real-time gravitational wave detection	10.6084/m9.figshare.26799172
Walker-Loud, André	2311431	Preparing LQCD calculations of multi-nucleons on heterogeneous architectures	10.6084/m9.figshare.26539786
Wang, Yang	2103958	A linear scaling ab initio approach to the electronic transport in disordered alloys	10.1184/R1/26763736.v1
Wang, Chuang	2417717	Cloud Infrastructure-Enabled Training for AI in Educational Research and Assessment	10.6084/m9.figshare.26800969
Wang, Guang	2411151, 2411152, 2411153	MobilityNet: A Trustworthy CI Emulation Tool for Cross-Domain Mobility Data Generation and Sharing towards Multidisciplinary Innovations	10.6084/m9.figshare.26520382.v1
Wang, Jason	2320147, 2320148	Cyberinfrastructure-Enabled Machine Learning for Understanding and Forecasting Space Weather	10.6084/m9.figshare.26337061
White, Joseph	2004954	Slurm Simulator Development: Balancing Speed, Accuracy, and Maintainability	10.6084/m9.figshare.26741788
Wu, Yinghui	2104007	Elements: Crowdsourced Materials Data Engine for Unpublished XRD Results	10.6084/m9.figshare.26798452
Xu, Zhihan	2311870	Portable Library for Homomorphic Encrypted Machine Learning on FPGA Accelerated Cloud Cyberinfrastructure	10.6084/m9.figshare.26540695
Yang, Xianfeng (Terry)	2234292	Simulation Platform for Evaluating Autonomous Vehicle Safety on Snowy and Icy Roads	10.6084/m9.figshare.26798920.v1
Yang, Yuxin	2209563	High-Performance Training of GNNs and TGNNs: Leveraging Parallel Computing Architectures and Hardware Accelerators	10.6084/m9.figshare.26540740
Yu, Hui-Chia	2311466	Elements: Open-Source Battery Electrode Simulation Toolkit using MFEM (BESFEM)	10.6084/m9.figshare.26530609.v1
Xu, Zhihan	2311870	Portable Library for Homomorphic Encrypted Machine Learning on FPGA Accelerated Cloud Cyberinfrastructure	10.6084/m9.figshare.26540695
Yang, Xianfeng (Terry)	2234292	Simulation Platform for Evaluating Autonomous Vehicle Safety on Snowy and Icy Roads	10.6084/m9.figshare.26798920.v1
Yang, Yuxin	2209563	High-Performance Training of GNNs and TGNNs: Leveraging Parallel Computing Architectures and Hardware Accelerators	10.6084/m9.figshare.26540740
Yu, Hui-Chia	2311466	Elements: Open-Source Battery Electrode Simulation Toolkit using MFEM (BESFEM)	10.6084/m9.figshare.26530609.v1
Zabih, Ramin	2311521	HTML on arXiv to support the blind and visually impaired	10.6084/m9.figshare.26540938
Zender, Charlie	2004993	New Convention to Encode Lossy Compression Metadata for Greener Dataset Storage	10.6084/m9.figshare.26516062
Zhang, Zhao	2401245	Diamond: Democratizing Large Neural Network Model Training for Science	10.6084/m9.figshare.26548312
Zhang, Liqing	2004751	CIWARS: CyberInfrastructure for Waterborne Antibiotic Resistance Risk Surveillance	MISSING
Zhang, Xiangliang	2321054	C2D - Cybertraining for Chemical Data scientists	10.6084/m9.figshare.26798662.v1
Zhang, Xiaodong	2310510	Research and Development of Open Source Software Libraries for Foundational Algorithms with Hardware Accelerations	10.6084/m9.figshare.26800852
Zhang, Zhe	2321069	Broadening Adoption of Cyberinfrastructure and Research Workforce Development for Disaster Management	10.6084/m9.figshare.26527999
Zhang, Zhao	2411294	Frameworks: hpcGPT: Enhancing Computing Center User Support with HPC-enriched Generative AI	10.6084/m9.figshare.26548312