2024 Cyberinfrastructure for U.S. NSF Major Facilities Workshop Report

Collaboration in Action

June 24, 2024

Cover. Designed by Christina Clark.
Photos on cover and throughout report by Angelic Rose Hubert, University of Notre Dame, and Nicole Virdone, CI Compass.

June 24, 2024
Version 1.0
CI Compass (https://ci-compass.org/)

Preferred Citation: Baldin, I., Brower, D., Butcher, D., Casey, R., Clark, C., Deelman, E., Flynn, B., Hasan, M., Kee, K., Livny, M., Mandal, A., Murillo, A., Nabrzyski, J., Pascucci, V., Petruzza, S., Romsos, C., Stanzione, D., Vahi, K., & Virdone, N. “2024 Cyberinfrastructure for U.S. NSF Major Facilities Workshop Report”, June 24, 2024. doi: 10.5281/zenodo.11372561.

Report Authors

- Ilya Baldin, Jefferson Lab
- Don Brower, University of Notre Dame
- David Butcher, Florida State University, NSF National High Magnetic Field Laboratory
- Rob Casey, NSF EarthScope Consortium
- Christina Clark, University of Notre Dame
- Ewa Deelman, University of Southern California
- Bob Flynn, Internet2
- Mahedi Hasan, Texas Tech University
- Kerk Kee, Texas Tech University
- Miron Livny, University of Wisconsin-Madison, PATh
- Anirban Mandal, RENCI
- Angela Murillo, Indiana University
- Jarek Nabrzyski, University of Notre Dame
- Valerio Pascucci, University of Utah
- Steve Petruzza, University of Utah
- Chris Romsos, University of Oregon, NSF Regional Class Research Vessel
- Dan Stanzione, University of Texas at Austin, Texas Advanced Computing Center
- Karan Vahi, University of Southern California
- Nicole Virdone, University of Southern California

Table of Contents

This project is supported by the U.S. National Science Foundation Office of Advanced Cyberinfrastructure in the Directorate for Computer Information Science under Grant #2127548. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agency.

Executive Summary

The 2024 Cyberinfrastructure for Major Facilities (CI4MFs) Workshop, organized by U.S. National Science Foundation (NSF) CI Compass [1], the NSF Cyberinfrastructure Center of Excellence, brought together cyberinfrastructure (CI) professionals from the NSF major and midscale research facilities along with participants from the broader CI ecosystem to discuss issues of critical importance to the success of these facilities. The workshop, which followed the series of events hosted by NSF CI Compass, was themed “Collaboration in Action” to underscore the need for collaboration when solving challenges faced by the CI community. As such, the workshop facilitated discussions via panels, break-out sessions, and small group conversations anchored in selected presentations from NSF and the community. CI Compass also took the opportunity to update the community on activities including the CI Compass Fellowship Program (CICF), now in its third year, on the work of the Cloud and FAIR Data Topical Working Groups focused on the use of clouds for Major Facilities (MFs) and approaches of making MF data FAIR (findable, accessible, interoperable, and reusable) [2], respectively, and on the research focused on building a community around CI.

Topics chosen for discussion at the meeting resulted from a community survey conducted by CI Compass in the Fall of 2023 and included: the approach to open science, combining data processing, movement, and storage, making MF data more accessible through visualization and analytics, and use of national CI. The workshop also included a highly interactive panel on lessons learned. Some of the workshop content was informational, some resulted in specific findings and recommendations, all meant to identify issues the community can work on further moving forward. As in the prior workshops, we encouraged them to create their calling cards, illuminating their successes, CI challenges, and thoughts about our community. The calling cards also serve as a community-building activity, virtually introducing participants to each other and enabling them to better connect in person.

The report sections are organized based on the workshop sessions and contain the findings, and recommendations if any. This summary includes only the highlights.

Key Findings:

- *Designing, building, and operating CI for MFs* is becoming more challenging as the pace of technical innovations is ever-increasing. This means that decisions that were made at design time often do not hold when the MF enters implementation and operational phases. Flexibility is key.
- *Cloud infrastructure* may be a solution for operations and data archival for some MFs but the implications of using commercial cloud providers are complex and a number of issues such as costs, efficient CI design, and workforce training need to be considered.
- *Data archiving i*s an important issue and there is a need to emphasize the importance of secure and cost-effective solutions.
- *Shared cyberinfrastructure* such as data storage can provide value across MFs.
- *Open science* considerations are growing in importance. MFs can lead the way in implementing and setting standards for open science.

Key Recommendations:

- *Designing, building, and operating CI for MFs:*
    - CI design should account for future changes at the various stages of the MF life cycle.
    - Science community expectation management should be a major goal when constructing new MF CI or updating existing MF CI.
- *Shared CI*: There is a need to explore potential benefits, risks, and funding models to support collaboration and outsourcing of critical functions.
- *Cloud infrastructure*: There is a need for collecting best practices on how best to re-architect the CI for MFs considering a move to the public cloud, to optimize costs and performance.
- *Data Archiving*: There is a need to assess long-term data preservation demands for not only MFs but, potentially, for a broader NSF community ecosystem and facilitate community discussions.
- *Open science:* To promote the sharing of data, there is a need to align incentives between researchers and facilities, provide documentation, and tutorials, and measure the impact that open data is having on the scientific landscape.
- *Open science:* The extra work involved in producing open data, documentation, and other materials needs to be recognized and incentivized by incorporating it into staff evaluations and promotions.

Overview and Goals

The goal of the CI4MF 2024 workshop was for MFs, members of the CI community, and NSF to share best practices, discuss opportunities, and brainstorm solutions to the challenges the community faces today and in the future. The theme of this workshop, “Collaboration in Action”, was chosen to emphasize the need for collaboration and for building a community to address both technical and social challenges faced by the broader MF CI community.

The workshop topics were solicited from the MF CI community during Fall 2023. Workshop topics included (a) MF Approach to Open Science: FAIR Data, Persistent Identifiers, Etc. (b) Use of National CI for MFs and Their Users, (c) Making MF Data More Accessible: Data Visualization and Analytics, (d) Coordinating and Combining Data Processing, Movements, and Storage, and (e) Approaches to CI Conceptualization and Design. The workshop was a continuation of prior meetings in 2017 [3], 2019 [4], and 2022 [5]. The workshop included a mix of talks, panels, and breakout sessions. The community was also invited to give lightning talks. Each session during the workshop was organized by a CI Compass member and a collaborator from an MF.

All presentations were recorded and made available online [6] and notes during the presentations and discussions were taken. This report summarizes the findings from the workshop and recommendations made by the participants.

Attendees of the 2024 Cyberinfrastructure for NSF Major Facilities workshop in Long Beach, California.

Workshop Structure and Activities

The workshop took place in Long Beach, CA, on January 17-18, 2024, with plenary and lightning talks broadcast virtually over Zoom. For a detailed agenda, please see Appendix C. The session speakers,

session hosts, and lightning talk presenters were all in-person attendees. The workshop talks, associated questions, and discussions were broadcast via Zoom.

Before the workshop, participants were asked to complete a Calling Card, where they could provide information about themselves and their facility, as well as get to know more about fellow attendees. These cards, which can be found in Appendix D, were streamed on the projector during breaks in the programming. An analysis of these cards can be found in the Calling Card Analysis section.

The workshop consisted of an update on CI Compass activities, one keynote by Katerina Antypas about the NSF NAIRR-Pilot effort, an important multi-agency effort in the US that aims to provide computational, data, software, and educational resources for artificial intelligence (AI) researchers and educators in the U.S, two invited talks by Michael Corn about the connection between cybersecurity and data management in support of open science and Mike Prince, who talked about the Antarctic Research Vessel (ARV) Project and its progress to build a new research vessel, two lightning talk sessions, two breakout session talks (*Making MF Data More Accessible: Data Visualization & Analytics*, and *Use of National CI for MFs and Their Users*), and three panels (*MF Approach to Open Science: FAIR Data, Persistent Identifiers, Etc.*, *Coordinating and Combining Data Processing, Movements, and Storage*, and *If I knew then what I know now, I would have never designed it this way: Approaches to CI conceptualization and design*). Each talk and session had a CI Compass chair. Additionally, each panel and breakout session had an "MF Buddy," that is, a CI practitioner from an MF who helped create guiding questions and moderate the sessions.

In addition to question and answer periods at the end of each session, in-person participants were able to enjoy ample informal networking and mixing opportunities during the three coffee breaks and during breakfast and lunch meal times each day. CI Compass also hosted a reception at the conclusion of Day 1 and a Social Coffee Hour at the conclusion of Day 2.

At the end of the workshop, in-person participants broke out into "tabletop discussions". Participants discussed current and future challenges and submitted an online form to CI Compass with their conclusions. Additionally, in-person and virtual attendees were sent an end-of-workshop survey where they could anonymously provide feedback to CI Compass.

All the workshop materials can be found at:
https://ci-compass.org/news-and-events/events/cyberinfrastructure-for-nsf-major-facilities-2024/

Update on CI Compass Activities

FAIR Data Topical Working Group

The FAIR Data Topical Working Group (TWG) was formed in August 2022 in response to CI Compass hearing a growing number of questions related to data management. The group meets monthly via Zoom and is open to all, including people who are not affiliated with MFs.

Members of the NSF CI Compass FAIR Data TWG were in attendance at the 2024 CI4MF workshop, including: Sarowar Hossain, Nicole Virdone, David Butcher, Shawna Sadler, Don Brower, Christina Clark, Jonah Kanner, and Angela Murillo.

During 2023, the group invited speakers from selected groups that work on FAIR data [2] to present their challenges and solutions, including the FAIR for Facilities and Instruments FAIROS RCN, the National Library of Medicine, and the NSF. The TWG organized a workshop titled "FAIR for Large Research Facilities" at the June 2024 ACM/IEEE Joint Conference on Digital Libraries [7].

The TWG also conducted a survey on FAIR data at MFs. The highlights are that facilities have done broad internal outreach on FAIR, and a majority have developed facility-wide plans for implementing FAIR practices. Having an understanding of leadership and people who are knowledgeable about FAIR were cited as very helpful. The big issues hindering adoption of FAIR practices were lack of time and money. Survey participants indicated that having more organizational support and more time would help improve FAIR adoption. For a more detailed analysis of the results, please see the group's report [8].

For 2024, the group plans to continue engaging with the FAIROS RCNs [9] and to invite additional speakers, especially for public webinars. We also plan to engage with ESIP [10], a group focused on data management for the Earth sciences.

To learn more about the FAIR Data TWG, visit https://ci-compass.org/about/working-groups/ or email contact@ci-compass.org.

CI Community Building

A community of practice (CoP) is "a group of people who share a common interest, profession, or passion and actively engage in collaborative learning and knowledge sharing" [11]. This definition describes the community of MFs well. This section of the CI4MF Workshop report summarizes three questions the Community Building Initiative team, especially the work by graduate students at Texas Tech University (TTU; Mahedi Hasan, Oluwabusayo Okunloye, Joni Litsey, Chaitra Kulkarni, and Madison Wedge) explored as part of the CI Compass's efforts in MF Community Building focused on the community of CI professionals. The findings are based on 23 interviews conducted by the TTU team via Zoom. Among the 23 interview participants, 13 are directors/managers, 4 project staff, 3 technical staff, and 3 scientists.

Interview Q1: Is there a sense of community among MFs?

In general, interview participants reported having a sense of MF community. Someone expressed that the sense of MF community is "when people know each other [and] share experiences with each other" (interview excerpt). However, it was also noticed that this sense of community was mainly reported by participants at the management level. It makes sense because these are the MF professionals who travel to conferences, network with other MF professionals outside of their facilities, and think about their MF/work beyond the context of their own MF/work. Furthermore, it was reported that most MFs work on very specific scientific problems, so they often feel siloed. That said, common problems (e.g., cloud, cybersecurity, long-term data archive strategies, the pandemic) can help cultivate a sense of community among MFs.

Interview Q2: Why is building a CoP for MFs important?

Participants responded to this question with a general notion of a tension between time and values. In other words, everyone is busy, and a community of practice needs to provide added value for others to want to participate in it. There are two general added values shared by the participants:

1. Learn from others who came before you. One participant shared, "If you talk about the build phase, there are other MFs out there that have been there for 20 years. And so you could ask them the question – *Well, what do you do with data once it's 10 years old or 20 years old? Do people still want it? How do you archive it?*" (interview excerpt).
2. You don't know when (or why) you need an MF community. Another participant noted, "You might not need it right away, but when you need it, you're gonna be happy when you do need it. And you don't know when you're gonna need it" (interview excerpt).

Kerk Kee, NSF CI Compass senior personnel and professor at Texas Tech University, gave a presentation on CI Community Building at the 2024 CI4MF workshop.

Interview Q3: What strategies can help build a CoP for MF professionals?

Ten strategies emerged from the analysis:

1. Identifying common interests/ challenges,
2. Promoting knowledge sharing,
3. Organizing guest sessions,
4. Providing networking opportunities,
5. Facilitating collaborative projects/initiatives,
6. Exploring new partnerships,
7. Establishing community guidelines by the community,

8. Creating an inclusive environment,
9. Maintaining feedback loops,
10. Regularly evaluating and iterating.

Some of these strategies reflect what CI Compass is already doing, and others can inform future efforts. A handout of infographics [12] was shared, which was based on a conference paper in 2022 from *CI4Resilience* [13].

To learn more about Community Building at CI Compass, email contact@ci-compass.org.

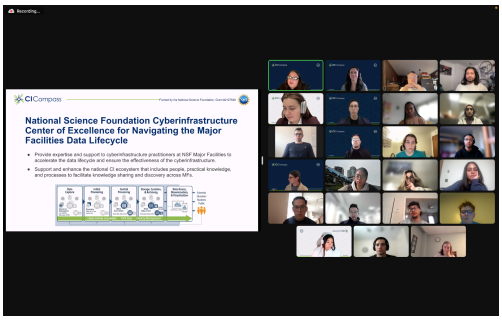
CI Compass Fellowship Program

Angela Murillo, director of CICF and professor at Indiana University - Indianapolis, presented about the fellowship program at 2024 CI4MF.

The CI Compass Fellowship Program (CICF) was created to broaden student participation in CI research, development, deployment, and operation and provides undergraduate student fellows the opportunity to: 1) learn about CI development and MFs, 2) develop CI-related skill sets important to the work of MFs, 3) engage with CI Compass and MF personnel and 4) participate in an optional/invited summer program to apply the skills learned for a particular MF project.

CICF consists of a Spring Program where fellows learn technical skills relevant to CI, learn about the importance and context of MFs, and research a specific MF to learn about its science mission, CI, and data assets. The Spring Program is virtual, free, and open to all undergraduate students attending a U.S. university or college, including community colleges. The Spring Program may be taken for course credit (depending on the student's institution requirements). Fellows are provided travel reimbursement to attend a professionally-relevant domestic conference. Additionally, university and college faculty can become involved by becoming CICF Faculty Mentors. Faculty Mentors are provided a stipend to help recruit potential fellows for the Spring Program, provide course credit or auditing options, hold check-ins with fellows, and provide feedback to the CICF team.

Fellows may apply for an optional, invited, and paid Summer Program to gain hands-on experience. In the summer of 2023, five student fellows were placed with the NSF National Center for Atmospheric Research (NSF NCAR) [14] / the NSF National Ecological Observatory Network (NSF NEON) [15] and two at the NSF Ocean Observatories Initiative (NSF OOI) [16] to work on a variety of projects.



A screenshot of the 2024 CICF program's online experience. The 2024 cohort included students from around the United States.

In 2024, CICF accepted 20 fellows in Spring 2024 from 17 institutions across 11 states into its third cohort. The program had 128 applicants from over 50 institutions. CICF 2024 Student Fellows will be placed at NSF NCAR/NSF NEON, NSF OOI, the NSF National High Magnetic Field Laboratory (NSF MagLab) [17], the NSF National Solar Observatory (NSF NSO) [18], the University of Notre Dame

Environmental Research Center (UNDERC) [19], and Globus Labs [20] at the University of Chicago for the 2024 Summer Program.

As the program continues to grow, there are various ways MFs can engage with CICF, including providing feedback on the program structure and content, participating as a guest speaker during the spring program, providing summer opportunities for students, and spreading the word to students, faculty, or others who may be interested.

To learn more about the program visit https://ci-compass.org/about/fellowships/ or email us at cicf@ci-compass.org.

Cloud Topical Working Group: Cloud Report

The cover image of the Cloud TWG's Cloud Report.

The CI Compass Cloud TWG was formed in 2020 in response to a survey of the MFs conducted by CI CoE Pilot (the precursor to CI Compass) on their data lifecycle (DLC), whereby the respondents identified the need for the community to get together and explore how cloud computing infrastructures can be used in MF operations. Topics of cost, vendor lock-in, and storage were among the most mentioned as being challenging for the MFs to navigate. The Cloud TWG was formed shortly after, consisting of staff from MFs (such as the NSF Seismological Facility for the Advancement of Geoscience (NSF SAGE)/ the NSF Geodetic Facility for the Advancement of Geoscience (NSF GAGE) [21], NSF Ocean Observatories Initiative (NSF OOI) [16], NSF IceCube Neutrino Observatory [22], NSF National Center for Atmospheric Research (NSF NCAR) [14]), cloud providers (such as Jetstream [23] and Cyverse [24]), personnel from the National Aeronautics and Space Administration (NASA) archives, cloud practitioners from Internet2 [25], and NSF CI Compass staff members. Over the past two years, this diverse group engaged in in-depth discussions about the various MF's, their needs, and how best to leverage the Cloud and other national CI supported by NSF.

As an outcome of these discussions, the Cloud TWG embarked on writing a technical report *NSF Major Facilities Cloud Use Cases and Considerations* [26] that was published in January 2024. The report explores the reasons and incentives for NSF MFs to leverage cloud resources for their operations. The cloud offers an extensive range of services, including data storage, archival, processing, and data access that offer gains in scalability and availability. However, those considerations must be balanced against other considerations, such as cost. The NSF supports over 25 MFs that serve as cornerstones for the scientific community, gathering colossal amounts of data each year. MFs have a number of resources they can leverage for their operations, including their on-premises infrastructures, as well as NSF-funded cloud (e.g. Jetstream 2 [23]), high-throughput computing systems (e.g. PATh [27]), and high-performance resources available (e.g. through ACCESS [28]). The report focuses on how the cloud can provide hosting and/or archiving data solutions to MFs. This is appealing to MFs because it will make their data more accessible to the wider community. Within the report are case studies from the NSF National Center for Atmospheric Research (NSF NCAR) [14], NSF IceCube Neutrino Observatory (NSF IceCube) [22], and the combined NSF Seismological Facility for the Advancement of Geoscience (NSF SAGE) and NSF Geodetic Facility for the Advancement of Geoscience (NSF GAGE) facilities, operated by the NSF

EarthScope Consortium [21]. These case studies offer insights into how other NSF MFs may approach and consider adopting cloud usage in different ways.

The Cloud TWG has considered the many ways that MFs may explore pivoting data storage and archival practices to the cloud and has published the findings. The report authors have invited readers to provide feedback and considerations to the TWG.

The report abstract says, “While [the cloud] introduces a myriad of complexities, its advantages are manifold. By viewing cloud adoption as a flexible, case-by-case decision rather than an absolute ‘all-or-nothing’ option, MFs can make informed choices about the level and extent of cloud adoption, which has the potential to significantly amplify their capabilities, enhance research quality, and ensure efficient resource utilization. This report encourages MFs to educate themselves on the fundamentals of cloud computing in order to better understand and leverage the cloud tools and computing model where they best fit MFs’ specific needs, optimizing benefits while effectively managing challenges and complexities.”

To learn more about the Cloud TWG, visit https://ci-compass.org/about/working-groups/ or email contact@ci-compass.org.

Keynote and Invited Talks

Keynote: Leveraging NSF's Cyberinfrastructure to Support Major Facilities Research

Speaker: Katerina Antypas (U.S. National Science Foundation (NSF))

Katerina Antypas, NSF, was the keynote speaker at the 2024 CI4MF Workshop.

Abstract: The NSF Office of Advanced Cyberinfrastructure (OAC) has long invested in cyberinfrastructure to enable major facility science throughout the entire data lifecycle, including data filtering, storage, transfer, analysis, curation, sharing, and publication. This talk will highlight OAC's investments in supporting major facility science through centers of excellence, computing, networking, cybersecurity, and data and software systems. It will also provide an update on OAC's latest initiatives including the National Artificial Intelligence (AI) Research Resource Pilot.

Invited Talk: Enabling Open Science and Data Sharing: Trust Provenance, and Data Integrity

Speaker: Michael Corn (NSF)

Michael Corn, NSF, was an invited speaker on the second day of the 2024 CI4MF workshop.

Abstract: Open Science is driven by knowledge discovery and innovation and fueled by the wide dissemination of scholarly publications and data. Information Assurance (inclusive of cybersecurity, data protections (including privacy), cyber risk management, and resilience) provides tools that support the practical implementation of FAIR principles. This talk explores how FAIR, Information Assurance and Research Security are related and why each domain needs to better recognize their shared concerns.

Invited Talk: Designing Cyberinfrastructure for the Antarctic Research Vessel

Speaker: Mike Prince (NSF/Antarctic Research Vessel [40])

Mike Prince, NSF and ARV, was an invited speaker for the second day of the 2024 CI4MF workshop.

Abstract: The NSF Antarctic Research Vessel (NSF ARV) Project involves the design and eventual construction and delivery of a new icebreaking research vessel to replace the current vessel, Nathaniel B. Palmer. The project is just finishing the preliminary design phase and Cyberinfrastructure is a key component of the design focus. As a research vessel, the ARV is a floating

laboratory and data gathering system operating in the harsh environments of Antarctica and the Southern Ocean. Early planning is critical to ensure that data from all the installed instrumentation and that brought on board by embarked scientists is captured, distributed, analyzed and archived. In addition, modern day ships systems such as propulsion, navigation, communications and life support all rely on a robust cyberinfrastructure. This talk will give an overview of the process and extent of the Cyberinfrastructure design for NSF ARV.

Panel: MF Approach to Open Science: FAIR Data, Persistent Identifiers, Etc.

Panel Discussion

Shawna Sadler gives her presentation during the MF Approach to Open Science panel.

This panel was moderated by CI Compass's Don Brower and Angela Murillo. Large facilities deal with a large number of complex datasets. This data is useful to both scholarship, policy-making, as well as daily life. This raises questions of how to manage this data to enable its reuse by researchers, the government, and the public. Specifically, what kinds of systems have proved useful and not useful? Are there any small steps that the community can take to improve interoperability between facilities and scientists, and make good use of our cyberinfrastructure?

Panelists:

- David Butcher (NSF National High Magnetic Field Laboratory (NSF MagLab) [17]): "MF Approach to Open Science: FAIR Data, Persistent Identifiers, and Engagement"
- Jonah Kanner (NSF Laser Interferometer Gravitational-Wave Observatory (NSF LIGO) [29]): "FAIR Public Data: The Last Mile"
- Monica Youngman (National Oceanic and Atmospheric Administration (NOAA)/National Centers for Environmental Information (NCEI) [30]): "Realizing FAIR Data and Open Science: Improving the Findability, Accessibility, Interoperability, and Reusability of NOAA Data"
- Amber Boehnlein (Thomas Jefferson National Accelerator Facility (JLab) [31]): "Open Science: A Perspective"
- Shawna Sadler (ORCID (Open Researcher and Contributor ID) [32]): "ORCID & Major Facilities"

The discussion covered current efforts at open data, the internal incremental changes needed to implement it, and the cultivation of a community using the data outside the facility. NSF LIGO has established an open data portal and created documentation and tutorials on how to get started with using the data. They have also run an annual summer course as a boot camp for researchers and interested citizens to get started. NSF MagLab has incorporated ORCID identifiers and DOIs into their workflows, which allows for assessing their user base and impact. The NOAA National Centers for Environmental Information is implementing an incremental effort to create an open knowledge graph for their environmental data that could potentially integrate with the NSF Open Knowledge Network (OKN) efforts in the future. Jefferson Lab is heading the U.S. Department of Energy (DOE) High Performance Data Facility award to standardize data management infrastructure [33]. ORCID is paying attention to many international and national efforts for standardizing and requiring persistent identifiers for researcher identification.

Key Findings

- Open data presents several opportunities for Major Facilities including showcasing what facilities can do and providing a way to be involved in the public conversation around the broader scientific questions the facility answers. Facility staff can also use the open data as a starting point without needing to involve the entire consortium used to collect the data if there is one.
    a. In some fields, especially in areas involving the climate or environment, open data can help address false narratives in the public.
    b. It is important for facility leadership to express the importance of open data and to set realistic expectations for progress.
- Building a community around a facility's open data requires making a pathway for outside researchers to follow. This requires documentation and tutorials.
    a. Provide guidelines for working with the data, but also recognize people can develop their own methods. This means there is a possibility people might make errors and reach unsupported conclusions with the data.
- Persistent Identifiers such as digital object identifiers (DOIs), ORCID, and others are critical for the identification of objects and tracking usage.
    a. There is a role for facilities to lead the way in implementing persistent identifiers (PIDs) and setting standards.
    b. For data, there is no clear granularity to assign identifiers to. It depends on the discipline and how the data is expected to be used. The ideal scope size is small enough to be meaningfully cited without being too big.
- Aligning incentives between researchers and facilities is critical since creating open data takes effort
    a. Citation and reuse are common goals in the scientific community, primarily due to their use in the promotion and evaluation of work.
    b. Funders and facilities want to measure the impact of open data on their communities
- Measurement is critical, not only at the end with download and citation counts but also over the entire data lifecycle.

Recommendations

- Open data efforts need facility leadership to recognize the importance of achieving results and setting expectations.
- Research infrastructures need to provide documentation and materials to help people understand and use your facility's data and get started. Incorporate feedback to improve the documentation.
- The community needs to recognize the extra work involved in producing open data, documentation, and other materials. Incentivize the work and incorporate it into staff evaluations and promotion.

To learn more about the FAIR Data TWG, visit https://ci-compass.org/about/working-groups/ or email contact@ci-compass.org.

Panel: Coordinating and Combining Data Processing, Movements, and Storage

Panel Discussion

Ilya Baldin presents the panel topic and speakers.

This panel was moderated by Ilya Baldin of CI Compass alum now at JLab, and Chris Romsos from the NSF Regional Class Research Vessel (RCRV). Major Facilities (MFs) use a wide variety of approaches to acquire, process, and store many types of data in support of their goals. A common thread however is the ever-increasing speed at which these processes take place and greater bandwidths and storage volumes required to support them. Simultaneously, advances in hardware and software allow for rethinking and repositioning of activities where some of the steps in these workflows take place. The general paradigm of the acquire-transport-store process gives way to more flexible pipelines where some of the intelligence and computation are migrating closer to the source of the data or into the network. The goal of the panel was to discuss how different facilities deal with orchestrating the processes supporting their data acquisition and processing workflows and where they see future trends taking them.

The panelists were asked to discuss some of the following issues with respect to their experience with their MF:

1. How does your facility today deal with orchestrating data movement, processing, storage (and if needed, dissemination) for the benefit of the users?
2. What enabling technologies do you see that will make this job easier in the near future?
3. Are there paradigm-shifting changes you see coming on the horizon for your users and how they handle their data movement, processing, and storage?
4. What challenges do you see going forward for them and for your facility?

The following presenters took part in the panel:

- Rob Casey, NSF EarthScope Consortium [21]
    - The presentation described the logic behind and the process of migrating Earthscope infrastructure to the cloud (AWS)
- Julio Ibarra, AmLight Express [34]
    - The presentation described express network paths connecting science facilities in South America and the Caribbean
- Chris Romsos, NSF RCRV [35]
    - The presentation covered how the scientific data is acquired aboard ships in ARF, the types of instruments typically used, and how the data is then processed
- Benedikt Riedel, NSF IceCube Neutrino Observatory [22]
    - The presentation covered data management at IceCube for multi-messenger astrophysics.

In the discussion with the audience, several themes emerged:

The financial implications of migrating science infrastructure to the cloud are difficult to predict and account for in many cases. There were many questions from the audience about the assumptions that went into the financial model for NSF EarthScope. Panelists have mentioned that overhead rules at different universities regarding cloud spending may differ, making spending forecasts even more difficult. There was a discussion with the audience regarding the hidden costs of cloud migration to the stakeholder science communities.

Ilya Baldin, JLab, and CI Compass alum, spoke to the audience during the Coordinating and Combining Data Processing, Storage, and Movement panel.

On a related note, the discussion touched on the decision processes and the costs for periodic refreshes for the on-premise (on-prem) MF CI.

Another point of discussion centered around differences in culture in the MF builder, operator, and science stakeholder communities and how to keep those aligned and manage expectations around MF operations. Scientists rarely fully understand the complexity of the CI in MFs or the costs both in time and money of keeping the infrastructure running, they are primarily concerned with the availability of the CI for their science needs.

Key Findings

- The implications of migrating and keeping MF CI in commercial cloud providers are complex and a number of issues may be considered:
    a. Costs, both direct as well as related to overhead rules for a specific organization must be carefully considered. Many organizations offer “free” power/space/cooling for on-prem infrastructure making financial modeling challenging.
    b. The migration must include updating the CI so it can take full advantage of cloud elasticity both for cost as well as performance reasons. A direct migration of an existing on-prem CI into the cloud without modifications isn’t likely to bear good results.
    c. Similar to (b), science communities should be prepared to operate their analyses in the cloud to help better manage the costs and reduce egress fees for MFs.
- The cultural differences between CI builders, operators, and science stakeholders must be taken into account when constructing new CI. Early socialization of ideas within those communities as well as ensuring mission/goal alignment. It is important to manage expectations with respect to the performance and availability of MF CI.

Recommendations

- The community needs best practices for cloud “FinOps,” or financial development operations, to help estimate the costs of possible migration and sustainment of MF infrastructure in the cloud.
- There is a need for best practices in how best to re-architect the CI for MFs considering a move to the public cloud, with the goal of optimizing costs and performance

- As more MFs consider moving to commercial clouds, science communities will need methods and best practices in operating their analyses in the cloud environment
- Science community expectation management should be a major goal when constructing new MF CI or updating existing MF CI.

Panel: If I knew then what I know now, I would have never designed it this way: Approaches to CI conceptualization and design

Panel Discussion

Jarek Nabrzyski introduces the panelists and concept for the discussion and panel.

This panel was moderated by Jarek Nabrzyski of NSF CI Compass, and Rob Casey of the NSF EarthScope Consortium [21]. Cyberinfrastructure (CI) is a critical component of the Major Facilities (MFs). Being very large investments in research, MFs go through several lifecycle stages. Significant effort goes into each of these stages: planning and design, construction, operations, and periodic enhancements. These different points in an MF lifecycle have significant implications on how the CI supporting the MF is conceptualized, designed, and evolved over time. The charge for this panel was to discuss (a) what important factors should be considered when designing the CI for a new MF, (b) how should the CI requirements from the scientific community be captured and integrated into the CI design, (c) what should be considered to offer flexibility in the design to accommodate future changes and technology evolutions, and (d) how should CI enhancements be designed once MFs go into operations and how MF CI should be evolving during operations. In this session, we heard from four panelists - representatives from MFs who have experienced CI evolution through different lifecycle stages - about lessons learned and best practices in CI conceptualization and design and pitfalls that should be avoided in the future. The presentations were followed by community discussions on the topic. Our panelists were:

- Robert Casey, software engineering manager IV, NSF EarthScope Consortium [21];
- Dan Stanzione, associate vice president for research (UT Austin) and executive director, Texas Advanced Computing Center (TACC), representing Leadership-Class Computing Facility (LCCF) [36];
- Sabine Botha, assistant research professor, Biodesign Beus CXFEL Lab and Department of Physics, representing the NSF Compact X-ray Free Electron Laser NSF (CXFEL) [37]; and
- Frank Wuerthwein, director of the San Diego Supercomputer Center (SDSC), Open Science Grid (OSG) Executive Director, representing the Large Hadron Collider (LHC) [38]

Robert Casey discussed design experiences in merging two major facilities (NSF SAGE and NSF GAGE) in the Cloud. He presented several lessons learned in cloud deployment and adoption for the Common Cloud Platform (CCP).

Dan Stanzione discussed experiences in the design of the LCCF, and some unexpected outcomes related to technology planning more than eight years in advance. The conclusions were that planning for CI so much in advance led to future design changes closer to the actual construction time, including support for artificial intelligence (AI), power envelope, changes to data center strategy from on-site to colocation

service agreements, etc. User requirements had to be adjusted as well because predictions so far out became obsolete later on.

Sabine Botha shared lessons learned from the NSF CXFEL perspective - the importance of redundancy in personnel knowledge and documentation, to not over promise to scientists, to scope and document CI and data retention policy, and to not assume that users will automatically use CI best practices.

Frank Wuerthwein emphasized the importance of time-invariant characteristics of the science enabled by an MF, and the importance of designing and tailoring the CI and data systems based on the data usage patterns dominant in the community served by the MF (e.g. data is immutable for the LHC community).

Panelists for the "If I knew then what I know now, I would have never designed it this way" discussion presented lessons learned throughout their experiences in the CI community. From Left, Dan Stanzione, TACC; Robert Casey, NSF SAGE and NSF GAGE; Frank Wuerthwein, LHC; and Sabine Botha, NSF CXFEL.

Key Findings

- Some of the key lessons learned during migration of data management operations to the Cloud (e.g. as in Earthscope) were: (a) it is important to identify the external factors like lease/license expiry dates in planning the timeline for migration, (b) selection of Cloud platform played a big role in the final system design, (c) the design has to be flexible in the spectrum from refactor to rehome, (d) prototyping early on had significant benefits, (e) one can't underestimate the importance in making the workforce ready for the transition.
- Participants agreed that coming up with precise requirements for a CI design is a complex problem and prediction is hard. An anecdote was presented on how the power requirements from vendors were significantly off for the LCCF design, and how AI was not a significant factor considered for its design. But, both factors changed significantly for the final design and construction. One approach to address this is to include realistic checks in the original design to verify whether the implementation is supporting the needs and also have an appropriate contingency fund to be able to adapt to technology changes.
- Participants discussed that there might be a tendency to overestimate future requirements during the CI design process, but that runs the risk of funding not being approved. However, on the other hand, if the design is scaled down, it may result in under-designing the system. Another related discussion point was about how to pivot a CI design during the operations phase of an MF while keeping current operations rolling.
- Another key finding was that large MF CI projects sometimes have a tendency to produce bespoke interfaces and tools that hide the functionality underneath, and the black-box nature of the design of these tools creates technology debt that has an impact on the workforce.

Based on the experiences of the panelists and the audience, this session came up with the following recommendations.

Recommendations

1. MF CI design should account for future changes, i.e. opportunity to measure current functionality and affected changes during operations should be baked into the final design of an MF; these can be addressed during the enhancement phase of the MF life cycle.
2. Requirements from the user community should be taken into consideration for CI design; but at the same time, contingencies must be planned in advance to respond to those fluctuations.
3. Research infrastructures need to document external factors (e.g. lease expiration date, license expiration date) when creating timelines for CI design and enhancements.
4. Research infrastructures need to select a platform before finalizing the system design for Cloud migration and start prototyping with that platform early on in the migration process. Projects need to be prepared to re-architect from scratch.
5. MFs need to document and educate their CI personnel about the time-invariant characteristics of the scientific endeavors carried out by the MF to avoid evaluating and implementing every new technology and solution.
6. To enable future development and modifications and reduce technical debt, the CI design for tools should be transparent about the functionalities.

Breakout Session: Making MF Data More Accessible: Data Visualization and Analytics

Breakout Session Discussion

Valerio Pascucci stands near the projection, near Dan Stanzione, as they lead the breakout session.

The Breakout Session was moderated by Valerio Pascucci of NSF CI Compass, and Dan Stanzione of TACC, and started by identifying which CI data visualization and analytics are needed and are preventing users from accomplishing their tasks. Three main use cases were discussed: (i) use of large Light Detection and Ranging (LIDAR) data and challenges in configuring remote servers for the task; (ii) working in an environment where the data rate is very high requires all components to be able to process the data at the same rate (e.g., data acquired at a beamline needs to be processed and then modify the beamline for the next acquisition); (iii) data from telescopes and how to minimize storage by compressing the video acquisition to few frames which contain the relevant data (i.e., when relevant phenomena occur). It was also highlighted how often it is preferred to develop custom visualization tools tailored to the particular needs of the specific project (i.e., NSF Laser Interferometer Gravitational-Wave Observatory (NSF LIGO)).

Another discussion centered around what resources would be desirable with additional benefits. Unanimous comments were made about investing in more software developers and CI experts. Finally, the group discussed data archiving strategies, including the use of tape archives and cloud storage.

Key Findings

- Visualization and analysis software stacks are hard to deploy and use.
- There is a need for shared expertise and workforce (“it is more useful to have an expert (even as a shared resource) to ask questions, rather than trying to read a book and learn about the Cloud”).
- Data archiving is an important issue and there is a need to emphasize the importance of secure and cost-effective solutions.

Breakout Session: Use of National CI for MFs and Their Users

Breakout Session Discussion

Jarek Nabrzyski stands at the podium with Miron Livny sitting at the panelists table during the discussion.


Peter Couvares, Caltech and NSF LIGO, participated in the breakout session discussion.

Moderated by Jarek Nabrzyski of NSF CI Compass, and Miron Livny of PATh, this session aimed to engage the MF community in a vital dialogue about how MFs and their users use the national CI. The general goal was to discuss the challenges and aspirations regarding national CI resources. The participants were invited to share their experiences, insights, and visions to collectively shape the future of computational support in scientific research. This session was an excellent opportunity for collaboration, learning, and influencing the trajectory of CI development in support of groundbreaking scientific endeavors.

The session started with a contextualization of the discussion, i.e. how major facilities, like NSF Laser Interferometer Gravitational-Wave Observatory (NSF LIGO) [29] and the NSF National Optical-Infrared Astronomy Research Laboration (NSF NOIRLab) [39], can effectively utilize National CI. It was suggested that by treating use cases as distinct yet interconnected challenges, facilities and funding bodies can more effectively address the specific needs of each scenario, leading to enhanced operational efficiency and improved data accessibility for researchers worldwide.

Discussion focused on efforts and collaborations on project-specific problems that required specialized solutions. It was emphasized that addressing obstacles like awareness, ease of transition, and the pace of technological updates can make national CI more accessible, responsive, and valuable to the scientific research community. The discussion centered on three separate topics: (i) long-term data storage/archival, (ii) debate about cloud versus national CI, and (iii) outsourcing of certain CI tasks to external groups.

Key Findings

- Long-term commitment: It is important to make a long-term commitment to archival efforts comparable to the Library of Congress, to ensure that valuable scientific data is preserved for future generations. Such commitment is crucial for projects that generate data with substantial long-term scientific value.
- Cost assessment: Determining what proportion of a project's total cost should be allocated to data archiving is a significant challenge. It requires balancing the immediate research needs against the long-term benefits of preserving the data.

- The intrinsic value of data: Every byte of data is precious, and the significant costs incurred in generating such data justify the efforts and investments in ensuring its longevity and accessibility.
- Technological advancements and costs: Continuing technological advancements will help drive down the costs of data storage, making long-term archiving more achievable. This optimism is rooted in historical trends where increased efficiency and capacity in computing and storage technologies have resulted in lower costs.
- There is a potential of collaborative approaches for data storage at scale.
- The complex considerations involved in utilizing cloud storage for long-term data preservation.
- CI Compass should form a working group to assess long-term data preservation needs for not only MFs but, potentially, for a broader NSF community ecosystem.

Other Workshop Sessions

Lightning Talks

Several members of the CI community gave lightning talks throughout both days of the 2024 CI4MF workshop. From Left: Jim Basney, Trusted CI; Shawna Sadler, ORCID; Michael Zentner, SGX3; and Renaine Julian, FSU, NSF FAIR Open Science Research Coordination Network.

The following community members gave lightning talks:

Day 1

- *ACCESS and Major Facilities*, by Benedikt Riedel (University of Wisconsin-Madison, NSF IceCube [22])
- *Building Private GPTs and LLMs for Research and Daily Productivity*, by Tyson Swetnam (University of Arizona, CyVerse [41])
- *Science Gateways and the SGX3 Center of Excellence*, by Mike Zentner (San Diego Supercomputer Center, SGX3 [42])

Day 2

- *ORCID Lightning Talk*, by Shawna Sadler (ORCID [32])
- *FAIR Instruments and Facilities: Increasing the Discoverability and Traceability of Research Facilities and Instruments through Persistent Identifiers*, by Renaine Julian (Florida State University, NSF FAIR Open Science, Research Coordination Network [43])
- *A Brief Update from TrustedCI*, by Jim Basney (NSF TrustedCI [44])
- *Low Latency Data and the DATA MESH*, by Jameson Rollins (Caltech/NSF Laser Interferometer Gravitational-Wave Observatory (NSF LIGO) [29])
- *World Data System*, by Meredith Goins (World Data System [45]; Meredith was unable to attend, but her slides were made available to participants).

Tabletop Discussions

The last session of CI4MFs was devoted to “tabletop” discussions where participants discussed questions posed by the organizers. The discussions were meant to inform the work of the CI Compass and help the community point out areas of need.

In this report, we summarize the answers to two questions: 1) What are the challenges you are facing in your work? and 2) Are there some topical working groups you would like to see CI Compass form?

The challenges can be classified into three main groups:

1. *Funding* as it relates to the timelines and timeliness of the federal budgets, which can delay the MF project plans and the lack of appropriate funding for managing ever-increasing CI and data management costs (which include making data FAIR).
2. *Workforce development*, which encompasses issues of training of the current and future workforce, as well as MF users, creating fulfilling career paths for MF personnel, and associated issues of personnel retention.
3. *Keeping up with technologies*, not only in terms of workforce training but also in the context of the need to perform CI upgrades, to identify the right solutions, and to perform the upgrades in such a way as to minimize their impact on the user community while staying within the project's budget.

CI4MFs participants also indicated a number of topics they would like to discuss in future community meetings and CI Compass working groups. These included: data management, with particular emphasis on streaming data and data archiving, use of shared infrastructure either national or shared between MFs, workforce development, and use of AI for MFs. The participants also indicated their continued interest in the current CI Compass Cloud Topical Working Group and wanted to share experiences in using AI for MFs.

Calling Cards Analysis

Prior to the CI4MF 2024 Workshop, participants were asked to complete a Calling Card that was used to help participants meet each other virtually before the workshop and to facilitate conversations about topics of mutual interest between participants during the workshop. The Calling Cards asked participants to briefly answer three questions:

- *Please tell other participants about your biggest recent CI (including workforce development) accomplishment.*
- *Please share some thoughts about what are the things that keep you up at night or opportunities you see for your MF/project (in the context of CI).*
- *Please share your ideas about topics you would like our CI community to discuss together (technical or social).*

Given that the responses to the second question are more actionable, this summary focuses on the synthesis of these ideas.

Participants submitted 48 Calling Cards. We utilized the thematic approach in the social sciences to analyze the prompt answers.

In this section, we report the main themes of challenges and opportunities.

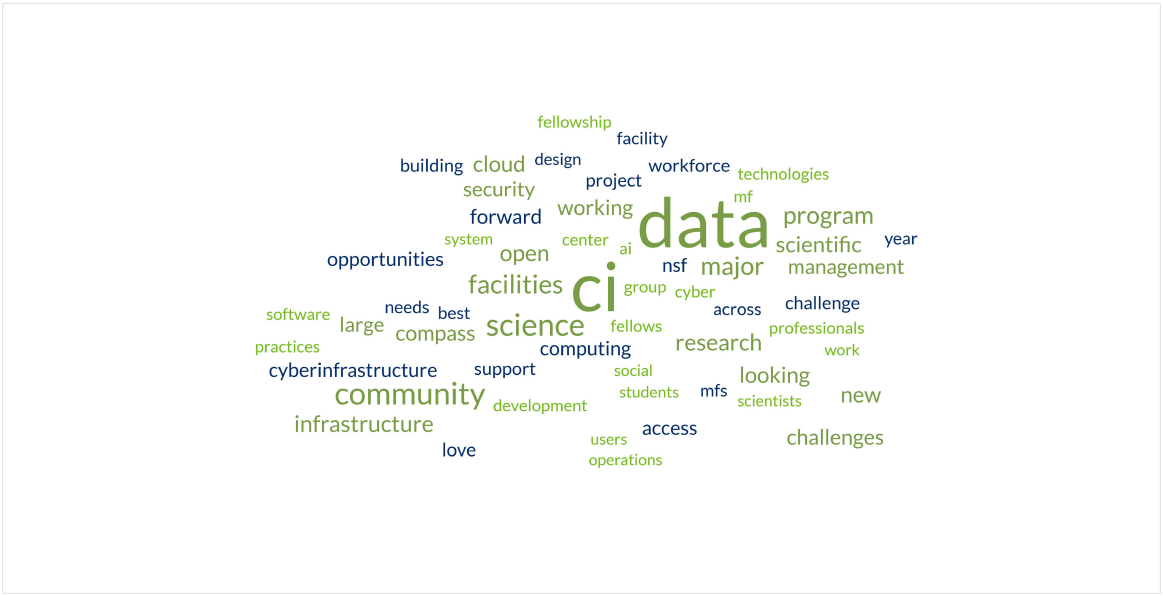


Figure 1: Word cloud of overall answers provided by the participants.

Challenges in NSF Major Facilities

The predominant challenges faced by MFs revolve around workforce development, communication, CI, managing big data, and community building. Workforce development was a consistent concern, encompassing both hiring and retaining skilled personnel, retraining existing staff, and addressing

diversity issues. Communication challenges included remote operating environments, coordinating data analysis plans, and bridging the digital divide. CI struggles involved designing adaptable systems, ensuring sustainability, defending against cybersecurity threats like ransomware, and transitioning to the cloud. Managing large and complex data sets on modest budgets was a recurring issue, with concerns about data integrity and accessibility. Building and sustaining communities, especially in diverse and unique organizational structures, presented ongoing difficulties.

We provide more details of these main themes of challenges below:

1. **Workforce Development and Diversity:**
    - Hiring and keeping skilled people is tough. Retraining the existing team is also a big challenge.
    - Addressing workforce turnover remains crucial, especially as the next phase of projects, such as the Large Hadron Collider (LHC) in 2029, demands more with fewer people.
    - Fostering diversity, both demographic and experiential, remains an ongoing challenge, requiring continuous efforts to bridge existing gaps within the community.

2. **Communication:**
    - Operating in remote and isolated environments, particularly at sea for extended periods, creates challenges in accessing communication and data systems.
    - Coordinating data analysis plans across institutions with varying expertise, and bridging the digital divide adds complexity to collaborative efforts.

3. **Cyberinfrastructure:**
    - The need for increased bandwidth, especially in the Antarctic region, calls for upgrading legacy capabilities to meet cybersecurity needs and support research effectively.
    - Designing cyberinfrastructure that accommodates an ever-changing user group poses a dynamic challenge.
    - Balancing the tension between building a comprehensive system and providing flexibility for diverse user needs remains an ongoing struggle.
    - Addressing cybersecurity threats, including ransomware, and uplifting existing 24/7 data repositories to the cloud requires constant attention and coordinated action.

4. **Managing Big Data:**
    - Handling big and complex data with limited budgets is a constant challenge. Archiving a lot of data every day is a specific worry.
    - Ensuring data integrity and accessibility, particularly in the face of hardware failures and cybersecurity breaches, is an ongoing concern.
    - Collaboration and agreement on conventions among various communities are critical for effective data management and sharing.

5. **Community Building:**
    - Making and keeping communities in diverse NSF Major Facilities is hard. Especially, figuring out ways for people to connect when working virtually is a big challenge.
    - Innovative communication practices and technologies are sought to reduce the feelings of isolation associated with virtual work.

- Challenges in constructing a community platform based on shared interests, research groups, and opportunities persist.

Figure 2: Word cloud of challenges and opportunities.

Opportunities for NSF Major Facilities

Despite the challenges, there are numerous opportunities for NSF MFs. These opportunities lie in extending research areas, particularly in social sciences, and bringing together social scientists and cyberinfrastructure developers for more effective technologies. Collaborative opportunities included expanding fellowship programs, involving more Minority-Serving Institutions (MSIs) and Historically Black Colleges and Universities (HBCUs), and fostering partnerships with MFs. Leveraging open-source composable computing, exploring AI's influence on CI, and pursuing advancements in cyber technology offer significant potential. Marketing opportunities can enhance visibility, and advancements in cloud technology provide chances to reimagine architecture for accessibility and scalability. Additionally, using AI for performance enhancements and embracing tech refresh opportunities can contribute to the future directions and applications of CI across scientific domains.

We provide more details of these main themes of opportunities below:

1. **Diversity:**
    - Opportunities lie in extending research areas, particularly by adding a new dimension to social science research.
    - Collaboration between social scientists and CI professionals can lead to more effective, ethical, and user-centered technologies.
    - Expanding fellowship programs, especially involving Minority-Serving Institutions (MSIs) and building partnerships with MFs, presents growth opportunities.

2. **Collaboration and Community Building:**

- Collaboratively find better ways to perform difficult tasks, such as CI Compass, the NSF National High Magnetic Field Laboratory (NSF MagLab), and others working on the complex problem of how to help associate metadata with instruments used in experiments.
- Building a community platform based on common interests, challenges, and opportunities, fostering existing relationships and knowledge sharing across the community.
- Opportunities include marketing initiatives to increase the visibility and reach of CI Compass within and outside the cyberinfrastructure community.

3. **Better Technologies:**
    - Upgrading technology, especially in data centers, can make things work better.
    - New ideas in cloud technology can change how we build things, making them more accessible and scalable.
    - Using advancements in cyber technology can improve the discovery, access, and interoperability of federal government data, addressing broader societal challenges.
    - Using open-source computing and looking at how AI can change things can have a big impact on future plans and uses.

In summary, while MFs face multifaceted challenges, embracing opportunities in diversity, collaboration/community building, and technology can pave the way for innovation and growth in both cyberinfrastructure and scientific research.

References

[1] Marketing Communications: Web and University of Notre Dame, “CI Compass,” *CI Compass*. Available: http://CI-compass.org. [Accessed: Mar. 21, 2024]

[2] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci Data*, vol. 3, p. 160018, Mar. 2016.

[3] “2017 NSF Large Facilities Cyberinfrastructure Workshop Report.” 2017.

[4] E. Deelman *et al.*, “2019 NSF workshop on Connecting Large Facilities and cyberinfrastructure,” 2019 NSF Workshop on Connecting Large Facilities and Cyberinfrastructure, 2020. doi: 10.25549/0ZBF-8M77. Available: https://facilitiesci.github.io/2019/2019_nsf_lf_ci_report.pdf

[5] E. Deelman *et al.*, “2022 cyberinfrastructure for NSF Major Facilities workshop report.” Zenodo, 2022. doi: 10.5281/ZENODO.6643901. Available: https://zenodo.org/record/6643901

[6] Marketing Communications: Web and University of Notre Dame, “Cyberinfrastructure for NSF Major Facilities 2024,” *CI Compass*. Available: https://ci-compass.org/news-and-events/events/cyberinfrastructure-for-nsf-major-facilities-2024/. [Accessed: Apr. 09, 2024]

[7] C. Clark, “CI Compass Topical Working Group leads efforts on FAIR data principles for NSF Major Facilities, to host workshop in June,” *CI Compass*. Available: https://ci-compass.org/news-and-events/news/ci-compass-topical-working-group-leads-efforts-on-fair-data-principles-for-nsf-major-facilities-to-host-workshop-in-june/. [Accessed: Mar. 21, 2024]

[8] A. P. Murillo, D. Brower, and C. F. Vardeman II, “Tech notes: Facilitating connections: The CI Compass FAIR Data Topical Working Group.” Zenodo, 2024. doi: 10.5281/ZENODO.10815117. Available: https://zenodo.org/doi/10.5281/zenodo.10815117

[9] “NSF Award Search: Advanced Search Results.” Available: https://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=7414&BooleanElement=Any&BooleanRef=Any&ActiveAwards=true. [Accessed: Mar. 21, 2024]

[10] PixelLighthouse, “Home - ESIP,” *PILI - ESIP*, Dec. 23, 2023. Available: https://www.esipfed.org/. [Accessed: Mar. 21, 2024]

[11] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, 1998.

[12] Creators Cassandra Hayes1 Alex Olshansky1 Ewa Deelman2 Kerk F. Kee1 Show affiliations 1. Texas Tech University 2. University of Southern California, *Leveraging Advanced Cyberinfrastructure and Developing Organizational Resilience for NSF Major Facilities in the Pandemic Era*. doi: 10.5281/zenodo.6347681. Available: https://zenodo.org/records/6347681

[13] C. Hayes, A. Olshansky, E. Deelman, and K. F. Kee, “‘No hitting the pause button’: Organizational resilience of scientific workplaces during the COVID-19 pandemic,” in *International Communication Association Conference*, May 27, 2022.

[14] “NSF national center for atmospheric research.” Available: https://ncar.ucar.edu/. [Accessed: Mar. 21, 2024]

[15] “National Ecological Observatory Network.” Available: https://www.neonscience.org/. [Accessed: Mar. 21, 2024]

[16] “Ocean observatories initiative,” *Ocean Observatories Initiative*, Mar. 20, 2024. Available: https://oceanobservatories.org/. [Accessed: Mar. 21, 2024]

[17] “National High Magnetic Field Laboratory.” Available: https://nationalmaglab.org/

[18] “NSF’s National Solar Observatory: Unlocking the mysteries of the Sun and its effects on Earth,” *NSO - National Solar Observatory*, Jan. 22, 2018. Available: https://nso.edu/. [Accessed: Mar. 21, 2024]

[19] Marketing Communications: Web and University of Notre Dame, “UNDERC,” *UNDERC*. Available: https://underc.nd.edu/. [Accessed: Mar. 21, 2024]

[20] “Globus labs.” Available: https://labs.globus.org/. [Accessed: Mar. 21, 2024]

[21] “EarthScope consortium,” *EarthScope Consortium*, Sep. 27, 2022. Available: https://www.earthscope.org/. [Accessed: Mar. 21, 2024]

[22] A. King-Klemperer and J. DeMerit, “IceCube Neutrino Observatory,” *IceCube*. Available: https://icecube.wisc.edu/. [Accessed: Mar. 21, 2024]

[23] D. Hancock, “Jetstream2,” 2023, Available: https://scholarworks.iu.edu/dspace/bitstream/2022/29422/2/NSF%20Award%202005506%20-%20Annual%20Report%20-%2010_1_2022%20-%209_30_2023.pdf
[24] “Home.” Available: https://cyverse.org/. [Accessed: Mar. 21, 2024]
[25] “Home,” *Internet2*, Dec. 05, 2019. Available: https://internet2.edu/. [Accessed: Mar. 21, 2024]
[26] G. B. Berriman *et al.*, “NSF Major Facilities cloud use cases and considerations.” Zenodo, 2024. doi: 10.5281/ZENODO.10481409. Available: https://zenodo.org/doi/10.5281/zenodo.10481409
[27] “Partnership to advance throughput computing,” *PATh*. Available: https://path-cc.io/. [Accessed: Mar. 21, 2024]
[28] “Home,” *Access*, Jul. 25, 2022. Available: https://access-ci.org/. [Accessed: Mar. 21, 2024]
[29] “LIGO Lab,” *LIGO Lab | Caltech*. Available: https://www.ligo.caltech.edu/. [Accessed: Apr. 09, 2024]
[30] “National centers for environmental information.” Available: https://www.ncei.noaa.gov/. [Accessed: Apr. 09, 2024]
[31] J. Grames, D. W. Higinbotham, and H. E. Montgomery, “Thomas Jefferson national accelerator facility,” *Nucl. Phys. News*, vol. 20, no. 3, pp. 6–13, Aug. 2010.
[32] “ORCID,” in *Encyclopedia of Library and Information Science, Fourth Edition*, CRC Press, 2017, pp. 3505–3509.
[33] “Jefferson lab to lead $300+ million high performance data facility hub,” *Thomas Jefferson National Accelerator Facility*. Available: https://www.jlab.org/news/releases/jefferson-lab-lead-300-million-high-performance-data-facility-hub. [Accessed: Apr. 09, 2024]
[34] “Home,” *Research Education Collaboration*. Available: https://www.amlight.net/. [Accessed: Apr. 09, 2024]
[35] “Regional class research vessel (RCRV),” *College of Earth, Ocean, and Atmospheric Sciences*, Aug. 08, 2019. Available: https://ceoas.oregonstate.edu/regional-class-research-vessel-rcrv. [Accessed: Apr. 09, 2024]
[36] “LCCF,” Jul. 25, 2019. Available: https://lccf.tacc.utexas.edu/. [Accessed: Apr. 09, 2024]
[37] “CXFEL,” *CXFEL*, Feb. 12, 2021. Available: https://biodesign.asu.edu/cxfel/. [Accessed: Apr. 09, 2024]
[38] “The large hadron collider,” *CERN*. Available: https://home.cern/science/accelerators/large-hadron-collider. [Accessed: Apr. 09, 2024]
[39] “NOIRLab / NSF’s National Optical-IR Astronomy Research Laboratory.” Available: https://noirlab.edu/public/. [Accessed: Apr. 09, 2024]
[40] Unites States Antarctic Program, “New Antarctic research vessel (ARV).” Available: https://future.usap.gov/arv/. [Accessed: Apr. 09, 2024]
[41] “Home.” Available: https://cyverse.org/. [Accessed: Apr. 09, 2024]
[42] I. Shunko, “sciencegateways - Home page.” Available: https://sciencegateways.org/. [Accessed: Apr. 09, 2024]
[43] “Findable accessible interoperable reusable open science research coordination networks (FAIROS RCN),” *NSF - National Science Foundation*. Available: https://new.nsf.gov/funding/opportunities/findable-accessible-interoperable-reusable-open. [Accessed: Apr. 09, 2024]
[44] “Trusted CI: The NSF cybersecurity center of Excellence,” *Trusted CI: The NSF Cybersecurity Center of Excellence*. Available: https://www.trustedci.org/. [Accessed: Apr. 09, 2024]
[45] “Home,” *World Data System*, Apr. 06, 2022. Available: https://worlddatasystem.org/. [Accessed: May 03, 2024]

Appendix A: Workshop Contributors

Planning Committee

- Christina Clark, University of Notre Dame
- Ciji Davis, University of Southern California
- Ewa Deelman, University of Southern California
- Anirban Mandal, RENCI
- Mats Rynge, University of Southern California
- Nicole Virdone, University of Southern California

Additional Contributions

Session leaders/hosts, breakout group leaders, note-takers, and other generous contributions from:

- Ilya Baldin, Jefferson Lab
- Don Brower, University of Notre Dame
- David Butcher, Florida State University, National MagLab
- Rob Casey, NSF EarthScope Consortium
- Christina Clark, University of Notre Dame
- Ciji Davis, University of Southern California
- Ewa Deelman, University of Southern California
- Bob Flynn, Internet2
- Sarowar Hossain, Indiana University
- Kerk Kee, Texas Tech University
- Miron Livny, University of Wisconsin-Madison, PATh
- Anirban Mandal, RENCI
- Angela Murillo, Indiana University
- Jarek Nabrzyski, University of Notre Dame
- Valerio Pascucci, University of Utah
- George Papadimitriou, University of Southern California
- Steve Petruzza, University of Utah
- Chris Romsos, University of Oregon, NSF Regional Class Research Vessel
- Mats Rynge, University of Southern California
- Dan Stanzione, University of Texas at Austin, Texas Advanced Computing Center
- Karan Vahi, University of Southern California
- Charles Vardeman, University of Notre Dame
- Nicole Virdone, University of Southern California
- Mahedi Hasan, Texas Tech University

Speakers and Panelists

- Katerina Antypas, NSF Office of Advanced Cyberinfrastructure
- Ilya Baldin, Jefferson Lab
- Jim Basney, TrustedCI

- Amber Boehnlein, Jefferson Lab
- Sabine Botha, Arizona State University, NSF Compact X-ray Free-Electron Laser Project
- Don Brower, University of Notre Dame
- David Butcher, NSF National High Magnetic Field Laboratory
- Rob Casey, NSF EarthScope Consortium
- Michael Corn, U.S. National Science Foundation
- Ewa Deelman, University of Southern California
- Bob Flynn, Internet2
- Julio Ibarra, Florida International University, AmLight, RAPTOR, EnviStor
- Renaine Julian, Florida State University, NSF FAIR Facilities and Instruments
- Jonah Kanner, NSF Laser Interferometer Gravitational-Wave Observatory
- Kerk Kee, Texas Tech University
- Miron Livny, Partnership to Advance Throughput Computing
- Angela Murillo, Indiana University
- Jarek Nabrzyski, University of Notre Dame
- Valerio Pascucci, University of Utah
- Mike Prince, U.S. National Science Foundation, NSF Academic Research Vessel
- Benedikt Riedel, NSF IceCube Neutrino Observatory
- Jameson Rollins, NSF Laser Interferometer Gravitational-Wave Observatory
- Chris Romsos, NSF Regional Class Research Vessel, U.S. Academic Research Fleet
- Shawna Sadler, ORCID
- Dan Stanzione, Texas Advanced Computing Center
- Tyson Swetnam, CyVerse
- Frank Wuerthwein, San Diego Supercomputer Center
- Monica Youngmann, NOAA/National Centers for Environmental Information
- Michael Zentner, SGX3 Center of Excellence for Science Gateways

Appendix B: Workshop Participants

In-person Participants

Name	Affiliation	Major/Mid-Scale Facility or Project
Stuart Anderson	Caltech	LIGO
Katerina Antypas	National Science Foundation	multiple
Ilya Baldin	Jefferson Lab	(previously) CI Compass
Jim Basney	NCSA	Trusted CI
Kenneth Bloom	University of Nebraska-Lincoln	U.S. CMS Operations Program
Amber Boehnlein	Jefferson Lab	HPDF
Adam Bolton	NSF's NOIRLab	NSF's NOIRLab
Sabine Botha	Arizona State University	Compact X-ray Free-Electron Laser Project (CXFEL)
Don Brower	University of Notre Dame	CI Compass
David Butcher	Florida State University	National MagLab
Rob Casey	EarthScope Consortium	SAGE and GAGE
Christina Clark	Notre Dame Research	CI Compass
Michael Corn	NSF	N/A
Peter Couvares	Caltech	LIGO Laboratory
Ciji Davis	ISI, USC	CI Compass
Ewa Deelman	University of Southern California	CI Compass
Bob Flynn	Internet2	N/A
Jeffrey Glatstein	Woods Hole Oceanographic Institution	Ocean Observatories Initiative
Oliver Gutsche	Fermilab	LHC/CMS
Mark Holl	Arizona State University, CXFEL Labs	Compact X-ray Free-Electron Laser Project (CXFEL)
Sarowar Hossain	Indiana University	CI Compass
Timothy Howard	National Science Foundation	USAP
Angelic Rose Hubert	University of Notre Dame	CRC
Julio Ibarra	Florida International University	AmLight, RAPTOR, EnviStor
Renaine Julian	Florida State University	NSF FAIR Facilities and Instruments
Jonah Kanner	Caltech	LIGO
Kerk Kee	Texas Tech University	CI Compass
Miron Livny	University of Wisconsin-Madison	PATh

Name	Affiliation	Major/Mid-Scale Facility or Project
William Miller	National Science Foundation	N/A
Angela Murillo	Indiana University-Indianapolis	CI Compass
Jarek Nabrzyski	University of Notre Dame	CI Compass
George Papadimitriou	University of Southern California	CI Compass
Valerio Pascucci	University of Utah	CI Compass
Sean Peisert	Berkeley Lab and UC Davis	Trusted CI
Steve Petruzza	University of Utah	CI Compass
Mike Prince	NSF/GEO/OPP/AIL	Antarctic Research Vessel (ARV)
Benedikt Riedel	UW-Madison	IceCube
Jameson Rollins	California Institute of Technology	LIGO
Chris Romsos	Oregon State University	RCRV (ARF)
Mats Rynge	USC / ISI	CI Compass
Shawna Sadler	ORCID	ORCID
Elizabeth Sexton-Kennedy	Fermilab	CMS
Kelli Shute	Indiana University Center for Applied Cybersecurity Research (CACR)	Trusted CI
Dan Stanzione	TACC/The University of Texas at Austin	Leadership Class Computing Facility - Horizon
Tyson Swetnam	University of Arizona	CyVerse
Kevin Thompson	NSF	N/A
Karan Vahi	USC Information Sciences Institute	CI Compass
Nicole Virdone	USC/ISI	CI Compass
Frank Wuerthwein	San Diego Supercomputer Center	SDSC
Monica Youngman	NOAA/National Centers for Environmental Information	Archiving NOAA's Environmental Data and Transformation to the Cloud
Michael Zentner	San Diego Supercomputer Center	SGX3 Center of Excellence for Science Gateways
Paweł Żuk	University of Southern California	CI Compass

Virtual Participants

Name	Affiliation	Major/Mid-Scale Facility or Project
Brad Barber	Texas A&M University	IODP
Bruce Berriman	Caltech/IPAC-NExScI	IPAC

Name	Affiliation	Major/Mid-Scale Facility or Project
Arnab Biswas	Comilla University	Mass Communication and Journalism
Meredith Goins	University of Tennessee	World Data System
Mahedi Hasan	Texas Tech University	CI Compass
Cassandra Hayes	Stephen F. Austin State University	N/A
Dawn Lenz	Battelle	National Ecological Observatory Network
Anirban Mandal	UNC Chapel Hill	CI Compass
Rajiv Mayani	USC	USC
Richard Oram	NSF	RIO
Tabitha Oyewole	Diffusion Innovation Lab	N/A
Alison Rockwell	NSF	Both
Jennifer Schopf	Texas Advanced Computing Center / UT Austin	LCCF / EPOC
Minhaz Uddin	Texas Tech University	Mass Communication

Appendix C: Agenda

Day 1 - January 17, 2024

7:00 AM → 8:00 AM **Breakfast**

8:00 AM → 8:15 AM **Workshop Welcome**

Speaker: Ewa Deelman (CI Compass/USC)

8:15 AM → 9:00 AM **Keynote: Leveraging NSF's Cyberinfrastructure to Support Major Facilities Research**

Speaker: Katerina Antypas (NSF Office of Advanced Cyberinfrastructure)

9:00 AM → 10:00 AM **Update on CI Compass Activities**

Speakers: Ewa Deelman (CI Compass/USC), Don Brower (CI Compass/UND), Kerk Kee (CI Compass/TTU), Angela Murillo (CI Compass/IUI), Bob Flynn (Internet2)

10:00 AM → 10:30 AM **Coffee Break**

10:30 AM → 12:00 PM **Panel: MF Approach to Open Science: FAIR Data, Persistent Identifiers, Etc.**

Speakers: Don Brower (CI Compass/UND), Angela Murillo (CI Compass/IUI), David Butcher (MagLab), Jonah Kanner (LIGO), Monica Youngman (NOAA), Amber Boehnlein (JLab), Shawna Sadler (ORCID)

12:00 PM → 1:45 PM **Lunch Break & Group Picture**

1:45 PM → 2:45 PM **Lightning Talks: 1**

Speakers: Benedikt Riedel (IceCube), Giri Prakash (ORNL), Mike Zetner (SDSC/SGX3), Tyson Swetnam (CyVerse)

2:45 PM → 3:00 PM **Visualization Demo**

Speaker: Valerio Pascucci (CI Compass/UU)

3:00 PM → 3:30 PM **Coffee Break**

3:30 PM → 4:45 PM **Breakout Sessions: 1**

3:30 PM **Making MF Data More Accessible: Data Visualization & Analytics**

Speakers: Dan Stanzione (TACC), Valerio Pascucci (CI Compass/U Utah)

3:30 PM **Use of National CI for MFs and Their Users**

Speakers: Jarek Nabrzyski (CI Compass/UND), Miron Livny (PATh)

4:45 PM → 5:05 PM **Breakout Session Reports Back**

5:05 PM → 7:00 PM **Reception w/ Cash Bar**

7:00 PM → 10:00 PM **Dinner (on your own; groups encouraged)**

Day 2 - January 18, 2024

7:00 AM → 8:00 AM **Breakfast**

8:00 AM → 8:05 AM **Day 2 Welcome**

Speaker: Ewa Deelman (CI Compass/USC)

8:05 AM → 8:30 AM **Invited Talk: Enabling Open Science and Data Sharing: Trust, Provenance, and Data Integrity**

Speaker: Michael Corn (NSF)

8:30 AM → 10:00 AM **Panel: Coordinating and Combining Data Processing, Movements, and Storage**

Speakers: Benedikt Riedel (IceCube), Chris Romsos (RCRV/ARF), David Mencin (Earthscope), Ilya Baldin (Jefferson Lab), Julio Ibarra (AmLight, RAPTOR, EnviStor)

10:00 AM → 10:30 AM **Coffee Break**

10:30 AM → 10:50 AM **Invited Talk: Designing Cyberinfrastructure for the Antarctic Research Vessel**

Speaker: Mike Prince (NSF/GEO/OPP/AIL)

10:50 AM → 11:50 AM **Lightning Talks: 2**

Speakers: Jim Basney, Meredith Goins (UT/WDS), Renaine Julian (NSF FAIR Facilities and Instruments), Shawna Sadler (ORCID)

11:50 AM → 1:15 PM **Lunch**

1:15 PM → 2:30 PM **Panel: If I knew then what I know now, I would have never designed it this way: Approaches to CI conceptualization and design**

Speakers: Anirban Mandal (CI Compass/UNC/RENCI), Dan Stanzione (TACC), Frank Wuerthwein (SDSC), Rob Casey (SAGE/GAGE), Sabine Botha (CXFEL)

2:30 PM → 2:55 PM **Tabletop Discussions**

Speaker: Ewa Deelman (CI Compass/USC)

2:55 PM → 3:00 PM **Workshop Closing Remarks**

Speaker: Ewa Deelman (CI Compass/USC)

3:00 PM → 4:00 PM **Social Coffee Hour**

Appendix D: Cyberinfrastructure Calling Cards

Find the calling cards of attendees of the 2024 Cyberinfrastructure for U.S. NSF Major Facilities Workshop on the following pages.

CI Accomplishment

Completed the deployment of the FABRIC MS-RI testbed (as now former FABRIC Project PI/Project Director). Looking forward to designing and deploying HPDF (High Performance Data Facility).

CI Challenges or Opportunities

Building any large CI system used by multiple stakeholder groups has a challenge of satisfying their usability requirements, creating a tension between building a 'cathedral' (one system to rule them all) vs. 'neighborhoods' (constructing multiple ways in the way the system can be used an extended both by builders and users).

Looking Ahead

- *Approaches to managing (acquiring, transporting, storing, curating, distributing) data in large CI systems.*
- *Finding, growing and retaining qualified staff*
- *How to create sustainable open CI software which includes (and perhaps is driven by) community contributions*

Ilya Baldin
HPDF Data Fabric Integration Architect
JLab

Jefferson Lab

https://www.jlab.org

CI Accomplishment

Trusted CI hosted the 2023 NSF Cybersecurity Summit at LBNL in October. Registration included 190 in-person and 131 virtual attendees, including 22 students and representatives of 12 Major Facilities. Workshop/training topics included Jupyter security, Zeek, AI/ML, secure programming, regulatory compliance, and physical security.

CI Challenges or Opportunities

Ransomware is an ongoing threat to our cyberinfrastructure, and multi-factor authentication (MFA) is a critical protection.

Looking Ahead

Let's work together to support and develop the cyberinfrastructure workforce, including cybersecurity professionals.

Jim Basney
Director, Trusted CI
NCSA/UIUC

TRUSTED CI
THE NSF CYBERSECURITY CENTER OF EXCELLENCE

https://trustedci.org/

CI Accomplishment

Member of working group preparing report on NSF Major Facilities Cloud Use Cases and Considerations

CI Challenges or Opportunities

How to manage large and complex data sets (including metadata) on modest budgets.

Looking Ahead

Use of cloud computing going forward. Open science.

Bruce Berriman
Senior Scientist
Caltech/IPAC

https://www.ipac.caltech.edu

CI Accomplishment

I'm studying to understand the adoption of CI from a social science perspective. My aim is to ensure smooth AI integration for optimal adoption outcomes.

CI Challenges or Opportunities

The opportunity is to add a new dimension to the area of social science research and challenge is to walk in a road less travelled by.

Looking Ahead

I am looking forward to seeing esteemed thinkers and understanding their way to move forward.

Arnab Biswas
Assistant Professor
Comilla University, Bangladesh

www.cou.ac.bd

CI Accomplishment

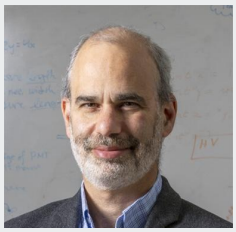
We store, transfer, and process many petabytes of data from the Compact Muon Solenoid at the Large Hadron Collider in Switzerland, and we've been doing it for 15 years. We are always looking at how to do this more efficiently and how to make life better for participating scientists.

CI Challenges or Opportunities

Data volumes will grow rapidly when an upgraded LHC turns on in 2029. How can we make sure keep up with it while keeping time to insight no bigger than it is now, within a flat-flat budget?

Looking Ahead

How can we continue to develop and retain an innovative workforce under constrained budgets? Are there paradigms that we can change before 2029 that can make a dramatic difference for our mission?



Ken Bloom
U.S. CMS Deputy Operations Manager
University of Nebraska-Lincoln

uscms.org

CI Accomplishment

Jefferson Lab is leading the development of the High Performance Data Facility(HPDF), partnering with LBNL.

CI Challenges or Opportunities

HPDF is will be a first in class facility for DOE. It is important for the design to take an ecosystem approach other CI activities

Looking Ahead

Data lifecycle management including data curation and data stewardship in the context of enabling scientific productivity

Amber Boehnlein
Computational Science and Technology Director
Jefferson Lab

jlab.org

CI Accomplishment

Data Storage, Curation and HTC analysis support for the NSF BioXFEL Science and Technology Center for 5 years. Currently we are in the process of developing the data analysis infrastructure for the CXLS/CXFEL X-ray sources currently being commissioned at ASU.

CI Challenges or Opportunities

The prospect of correctly archiving terabytes of data per day at kilohertz collection rates keeps me up at night. Coordinating data analysis plans across institutions and applications with varying expertises has been an ongoing challenge.

Looking Ahead

I am looking forward to networking with other scientists that have faced and conquered similar challenges to what we are currently encountering and learning from them.

Sabine Botha
Assistant Research Professor
Beus CXFEL Laboratory,
Arizona State University

https://biodesign.asu.edu/cxfel/

CI Accomplishment

My biggest recent success has been developing the curriculum for the CICF program. We want the fellows to have a broad understanding of CI and the data lifecycle to prepare them for internships and jobs.

CI Challenges or Opportunities

So much with data management depends on communities—not just disciplines but also scientists, engineers, and managers—having conversations and agreeing on conventions. I'm hopeful that we are at a phase transition to more open systems and sharing.

Looking Ahead

I would like to discuss fruitful next steps for making data and metadata more open, and for how to leverage open science to improve society.

Don Brower
CI Compass/Notre Dame

www.ci-compass.org

CI Accomplishment

I have provided leadership to software, policy, and work culture development with an incredible staff, leading to achieved milestones including complete data center migration to AWS, a maturing identity management system, and cloud-enabled data flow operations. I have also led this past year in hiring and enabling a half-dozen new developers.

CI Challenges or Opportunities

Cyberinfrastructure uplift to the cloud for existing 24/7 data repositories is a difficult task that requires constant attention to coordinated action as well as time and attention to the needs of technical staff. Our primary opportunity for CI in the cloud is to reimagine our architecture and pursue bold visions of accessibility and scalability not before possible.

Looking Ahead

I would like the CI community to devote more time to discussing:

- *Federated identity and access control between data repositories*
- *Approaches to quickly bringing a community around new standards*
- *Bringing scalable computing to the large data archive*
- *Meeting the needs for citability and reproducibility of large data sets*

Rob Casey
Software Engineering Manager
EarthScope Consortium
(GAGE/SAGE)

EarthScope Consortium

https://earthscope.org

CI Accomplishment

Being able to collaborate with communications professionals and top researchers at NSF Major Facilities like NCAR, NEON, OOI, and CEOAS/RCRV to tell the stories of how the engagements with CI Compass have a real-world impact on scientific discovery has been a big accomplishment. These impactful stories are also being driven by the CI Compass Fellows, who complete the CICF program and report back to us how much they have learned and the connections we have been able to help them make.

CI Challenges or Opportunities

I would like to continue to learn about and understand the organizational communication structures at play inside of the diverse NSF Major Facilities. Each facility is unique in both its mission and its organization.

Looking Ahead

I would love to speak with you about the impact meetings like CI4MF make in your corner of the cyberinfrastructure community. As a non-computer scientist, I would love to learn more about how CI impacts the science being done at your research facility.

Christina Clark
Communications Lead
CI Compass, Notre Dame Research

CI Compass
ci-compass.org

CI Accomplishment

It's been an absolute pleasure assisting Nicky and Angela on the CI Compass Fellowship program. It's inspiring to see such ambitious young scientist and it's great to be part of their professional growth.

CI Challenges or Opportunities

Possibly looking into marketing opportunities that would further our reach inside and outside of our CI community. It would be advantageous to the community and project to have a brighter light illuminate the CI Compass project.

Looking Ahead

How to engage the youth to increase interest in CI.
Engage the community with more social/fun ways to come together and network with one another.

Ciji Davis
Project Specialist
University of Southern California, Information Sciences Institute

https://www.isi.edu/

CI Accomplishment

2023 has been a busy year, but I think the biggest success for CI Compass has been the Student Fellows program. It is great to see undergraduate students learn about CI and the Major Facilities. Many also enjoyed their Summer internships at NCAR, NEON, and OOI. I look forward to meeting the class of 2024-- great job by the CI Compass CICF team!

CI Challenges or Opportunities

CI Compass has been working with MagLab and others to understand how to help associate metadata with instruments used in experiments. This is a very complex problem and I hope that we can come together as a community to make progress on it.

Looking Ahead

I see a lot of common CI issues across NSF Major Facilities and DoE and NOAA-funded projects in the US and other international efforts. I would love for CI Compass to play a bigger role in bringing these communities together. Ah, and I would love to visit facilities in Antarctica -- for a short while :)

Ewa Deelman
Research Professor
University of Southern California

https://ci-compass.org

CI Accomplishment

I contributed to the cloud working group's NSF Major Facilities Cloud Use Cases and Considerations *white paper.*

CI Challenges or Opportunities

A colleague and I are working to determine a metric for measuring the sustainability of computing choices.

Looking Ahead

How to develop incentives for including sustainability into not only TCO of CI, but in decision making about CI choices.

Bob Flynn
PM, Cloud Infrastructure & Platform Services
Internet2

Internet2.edu

CI Accomplishment

1 - Implementation of a "compute in place" environment using Jupyter Hub - further democratizing access to large data sets.
2 - Searchable library of digital stills and hd video with preview ability.
3 - Five Nines (99.999%) system availability for past project year.
4 - Real-time data plotting for cable delivered data.

CI Challenges or Opportunities

C1 - Loss of data through hardware failure or cybersecurity breach.
C2 - Workforce turnover.

O1 - Performance enhancements through tech refresh of the OOI data center

Looking Ahead

1 - Measuring data performance - how often is it accessed, how is it being used, how much has been downloaded and/or what is its quality?
2 - Are we presenting our data in a way that is best consumed by users? If not, how do we pivot to do so?

Jeffrey Glatstein
Senior Manager of Cyberinfrastructure
Ocean Observatories Initiative/Woods Hole Oceanographic Institution

oceanobservatories.org

CI Accomplishment

WDS's International Technology Office in Victoria, Canada served as co-chair to the Global Open Research Commons (GORC) International Model Working Group which released the GORC International Commons Model, helping to define the essential elements of commons https://doi.org/10.15497/RDA00097

CI Challenges or Opportunities

Sustainability of trustworthy scientific data in repositories while considering security, partnerships, staffing, and funding in both "normal" times and in times of crisis.

Looking Ahead

Opportunities to connect build collaborations and partnerships. Science and thus CI, can not be done alone as it takes all of us. WDS is always looking for ways to build bridges between its members and partners.

Meredith P. Goins
Executive Director
ISC's World Data System

worlddatasystem.org

CI Accomplishment

Extracting science from LHC proton-proton collisions requires a sophisticated and distributed computing infrastructure, flexible enough for the needs of 3000 individual researchers. From operating sites at national labs and universities to interconnecting them seamlessly amongst each other and the world, my teams bridge the gap between the common physicist and the expert software and computing infrastructure

CI Challenges or Opportunities

The next phase of LHC starting in 2029 will require to do much more with less people. How can we enable science on a scale that is orders of magnitude higher than today?
How can we be an effective partner in the global scientific computing infrastructure and not rely on home-grown solutions and technologies?

Looking Ahead

People are the key to success in scientific computing. How can we attract and retain talent? How can we provide meaningful career paths?

Oliver Gutsche
U.S. CMS Software & Computing Operations Program Manager
Fermilab - Computer Science and AI Directorate

uscms-software-and-computing

computing.fnal.gov

CI Accomplishment

Working on documenting best practices for CI adoption, implementation, and community building, and sharing these best practices with interested stakeholders.

CI Challenges or Opportunities

Challenges are the building a community platform based on interested, research group and opportunities.

Looking Ahead

Once again, looking forward to hear from CI Professionals and other personals towards developing a true community.

Mahedi Hasan
Research Assistant/Ph.D. Student
CI Compass/Texas Tech U

www.ci-compass.org

CI Accomplishment

Published a paper on the professionalization of CI and presented a paper on women professionals in CI at the National Communication Association Convention, along with co-authors Kerk Kee and Chaitra Kulkarni

CI Challenges or Opportunities

Security of CI and data, accessibility of CI and related tools and findings/data

Looking Ahead

How to further engage arts & humanities researchers and professionals with CI and MF-related projects

Cassandra Hayes
Assistant Professor
Stephen F. Austin State University

https://www.sfasu.edu/

CI Accomplishment

Implementation of the ASU Compact X-ray Light Source (CXLS) facility and light source with supporting Science DMZ with ASU Research Computing for facility and instrument commissioning. Currently we are in the process of developing network infrastructure upgrades for CXLS and are building a new compact XFEL at ASU.

CI Challenges or Opportunities

As an NSF Mid-Scale Reserach Infrastructure-2 project (CXFEL) we are preparing to deploy our network, data storage, and computational infrastructure for instrument control, data acquisition, and analysis. Having defined the sensor and actuator network, data types, and rates we are updating the hardware and software acquisition architecture.

Looking Ahead

My current objective is to engage and learn from this body of experts and community of practice and to bring new knowledge of best practices back to the CXFEL project prior to implementing critical procurements.

Mark Holl
Associate Research Professor
Chief Engineer - CXFEL Labs
Beus CXFEL Laboratory,
Arizona State University

ASU Biodesign Institute
Arizona State University

CXFEL Labs

https://biodesign.asu.edu/cxfel/

CI Accomplishment

For the 3rd year of CICF we received 120+ applications. We will be starting the 3rd year on 23rd January with 20 fellows.

CI Challenges or Opportunities

I would love to get more MFs involved so that fellows get more summer opportunities, guest speakers and networking opportunities. Enrich the CICF curriculum even more, and help the fellows as much as I can.

Looking Ahead

I think I would like to know more about concurrent CI issues, love input from MFs to our fellowship program also looking forward to connect with MFs.

Sarowar Hossain
Teaching Assistant
CI Compass Student Fellowship

INDIANA UNIVERSITY

CICompass/CICF

CI Accomplishment

Integrating the federal identity management solution within the U.S. Antarctic Program (USAP) for CI at Antarctic research stations.

CI Challenges or Opportunities

- *Provide increased bandwidth to support research in the Antarctic region.*
- *Work with science programs to upgrade legacy capabilities to meet current cybersecurity needs.*

Looking Ahead

Integration of university/institution identity management solutions with USAP identity management capability.

Tim Howard
USAP Information Security Manager
NSF Office of Polar Programs

https://www.usap.gov/

CI Accomplishment

I am a member of a project team that is working towards the advancement of the adoption of persistent identifiers for research facilities and scientific instruments.

CI Challenges or Opportunities

Finding ways to connect researchers, resources, and information in an effort to advance open science and open scholarship.

Looking Ahead

Open science, connected science, and scientific access/discovery

Renaine Julian
Director, STEM Libraries
Florida State University Libraries
co-PI FAIR Facilities and Instruments NSF

FLORIDA STATE UNIVERSITY
LIBRARIES

lib.fsu.edu

CI Accomplishment

Our GW Open Science Center (GWOSC) now hosts hybrid workshops with 15 locations and thousands of participants.

CI Challenges or Opportunities

New students are constantly struggling to learn how existing software works. How do we make it easier to get up this learning curve?

Looking Ahead

I hope we can share or create rubrics for measuring the success in applying FAIR principles to public data.

Jonah Kanner
Senior Scientist
LIGO Lab, Caltech

GWOSC

https://gwosc.org/

CI Accomplishment

My biggest CI accomplishment is working as a social and organizational scientist of cyberinfrastructure over the last decade or so, documenting best practices for CI adoption, implementation, and community building, and sharing these best practices with interested stakeholders.

CI Challenges or Opportunities

The biggest opportunity will have to include exploring how AI can influence the future directions of CI and its applications in various scientific domains. This is very exciting!

Looking Ahead

I am looking forward to seeing more productive conversations between CI and MF professionals towards a more vibrant community of CI4MF!

Kerk F, Kee
Senior Personnel/Professor
CI Compass/Texas Tech U

www.ci-compass.org

CI Accomplishment

NEON CI in the cloud - accomplishments in data access, infrastructure management, resource usage

CI Challenges or Opportunities

Data and infrastructure integrity. Flexibility and efficiency in data access. Role of AI.

Looking Ahead

Aligning strengths and assets across the community to foster innovation in CI for science

Dawn Lenz
CI Developer
National Ecological Observatory Network

neonscience.org

CI Accomplishment

Currently leading a group research project to explore how virtual collaboration can support the CI community.

CI Challenges or Opportunities

Continuing to build on existing relationships and foster sustainability in knowledge sharing across the community.

Looking Ahead

Identifying opinion leaders within the CI compass community who can model best practices and share knowledge with others throughout the community.

Joni Litsey
PhD Candidate
Research Assistant
Texas Tech University

CI Compass
ci-compass.org

CI Accomplishment

We enabled a NRAO team to process a dataset that was obtained a decade ago using GPU capacity provided by 12 institutions

In 2023 we served more than 460 users from 100 institutions consuming capacity provided by 60 institutions at a sustained rate of 5.18 jobs per second.

CI Challenges or Opportunities

Position CI as an enabler - serving as the Local and Community - of Translational Computer Science

Looking Ahead

Extend the reach of our software technologies and services to additional (we have 4 so far!) MFs

Miron Livny

Profesor
UW-Madison
PATh Project
OSG Consortium

https://path-cc.io/
https://osg-htc.org/

CI Accomplishment

Learning about the CI on research vessels. Particularly satisfying was the CI Compass work on gleaning CI requirements for the planned Antarctic Research Vessel in its early design phase.

CI Challenges or Opportunities

I would say that the two primary CI challenges/opportunities faced by major and mid-scale facilities are: (1) how to enable support for Open Science, and (2) how to leverage new advances in AI technologies to advance the science delivered by an MF and to improve the efficiencies of the CI operations within an MF.

Looking Ahead

The CI community should discuss different opportunities to train the workforce on Open Science best practices - both for the users and CI developers.

Anirban Mandal
Co-Principal Investigator
and CI Compass
Associate Director
CI Compass/UNC/RENCI

www.ci-compass.org

CI Accomplishment

We successfully migrated one of our large on-prem facilities to a pure cloud environment.

CI Challenges or Opportunities

Workforce development (both hiring and retaining) remain our a significant challenge.
Retraining our current workforce, both in technology and culture, remains a challenge.

Looking Ahead

Workforce development.
Educating new users (they too struggle with the change),

David Mencin
Vice President (CTO) - Data Services
EarthScope Consortium

EarthScope Consortium

www.earthscope.org

CI Accomplishment

We are in our third year of the CI Compass Fellowship Program, and each year, the program has a broader impact. The interest was phenomenal this year, with over 120 student applications. It is wonderful to see the fellows learn about MFs and CI and apply their skills to summer projects with our MF collaborators.

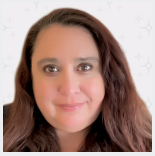
CI Challenges or Opportunities

I would love to continue expanding our fellowship program, including reaching more MSIs and HSIs. I would also love to expand our collaboration with MFs and have more opportunities for fellows to work with MFs during the summer program.

Looking Ahead

I look forward to continuing to refine the CI Compass Fellowship program curriculum and structure, engaging with more students, and building more collaborations with our MF partners.

Please reach out if you're interested in our fellowship program; we would love your feedback!



Angela Murillo
Co-Principal Investigator and CI Compass Student Fellowship Director

CI Compass/IU

www.ci-compass.org

CI Accomplishment

My biggest recent success has been the decennial workshop on the Future of Computing. The workshop has brought together CI leaders from the USA and Europe to discuss challenges and opportunities for CI in the next decade.

CI Challenges or Opportunities

We have a great opportunity and a need to bring social scientists to CI collaborations. By working together, social scientists and cyberinfrastructure developers can create more effective, ethical, and user-centered technologies that better serve the needs of various communities and MFs, and society at large.

Looking Ahead

I would like to fully understand the challenges MF face, including integrating new systems with existing legacy technologies, managing and storing massive data volumes, ensuring robust security and privacy, achieving interoperability among diverse systems, and maintaining scalability to support growing demands, as well as workforce development

Jarek Nabrzyski
Director, Center for Research Computing
University of Notre Dame
CI Compass/Notre Dame

www.ci-compass.org

CRC CENTER FOR RESEARCH COMPUTING

crc.nd.edu

CI Accomplishment

Just got a conference paper accepted on strategies for 'virtual Collaborations during the Pandemic' along with lead author, Kerk Kee and co-author Mahedi Hasan.

CI Challenges or Opportunities

What sort of innovative communication practices and technology (ies) can help reduce the feelings of isolation that is associated with virtual work?

Looking Ahead

Mental health and CI professionals
Socialization and community building

Oluwabusayo Okunloye
Graduate Student
Texas Tech University

CI Compass
ci-compass.org

CI Accomplishment

In 2023 we demonstrated how to achieve equity in access and use for data analysis and visualization of large scale scientific data. Practical examples include using gigabytes to petabytes of climate modeling and materials science data with no special local resources: https://nationalsciencedatafabric.org/dashboards.html .

CI Challenges or Opportunities

Empowering scientific communities in building, based on their needs, domain-specific CI solutions with scalabile tools and components that interoperate with the national cyberinfrastructure.

Looking Ahead

I am looking forward to the opportunity of expound partnerships with scientific communities to better understand their needs and priorities and provide support for their training programs and scientific use cases.

2024 Cyberinfrastructure for NSF Major Facilities Workshop

Valerio Pascucci
Professor
University of Utah

National Science Data Fabric

nationalsciencedatafabric.org

CI Accomplishment

1. Working with staff from design/construction teams for the RCRVs (OSU), CCRV (SIO), OOI (OSU & WHOI), and ARV (USAP) to build security into new and refreshed academic maritime ships and underwater vehicles by design.
2. Learning more about the CI challenges faced by traditionally underserved research institutions.

CI Challenges or Opportunities

Challenges:
Flat Major Facility budgets and increased cybersecurity needs/costs.
Recruiting skilled CI security workers.
Research security regulations.

Looking Ahead

How do we convince sponsors of the need to increase budgets to adequately fund necessary CI operations?
How do we hire and retain talented CI security staff?

2024 Cyberinfrastructure for NSF Major Facilities Workshop

Sean Peisert
Deputy Director, Trusted CI
Senior Scientist, Berkeley Lab
Adjunct Professor, UC Davis

https://trustedci.org
https://www.cs.ucdavis.edu/~peisert
https://secpriv.lbl.gov

CI Accomplishment

Supporting major facilities in their data management and visualization challenges.

CI Challenges or Opportunities

Finding common requirements in existing cyberinfrastructure to minimize replication of mature solutions.

Looking Ahead

Looking forward to understanding interoperability and scalability challenges in current MFs and how those could be addressed.

Steve Petruzza
Research Associate
University of Utah
Assistant Professor
Utah State University

www.ci-compass.org

National Science Data Fabric

CI Accomplishment

End-to-end data management for complex sensors deployed worldwide, successful operation of data center with 5+ petabytes of atmospheric data collected over 30 years. Providing advanced data discovery and continuous improvement of various user CI resources.

CI Challenges or Opportunities

Enabling seamless user experience by providing integrated data, computing, and software capabilities to facilitate cross-domain research.

Looking Ahead

Enabling a Data Workbench for scientists to perform long-term climate data analysis. AI adoption into Data Center operations to enhance efficiency, decision-making, and overall performance improvement.

Giri Prakash
Chief Data and Computing Officer

Atmospheric Radiation Measurement Facility, Oak Ridge National Laboratory

https://www.arm.gov/

CI Accomplishment

The Antarctic Research Vessel (ARV) project successfully completed the Preliminary Design Review and internal NSF reviews and received approval to move to the Final Design Phase. Cyber Infrastructure and Cyber Security plans are an integral part of the ARV design process and were reviewed favorably with useful recommendations for the next steps.

CI Challenges or Opportunities

Designing Cyber Infrastructure that will support an every changing group of users operating in a very remote and often hostile environment (i.e. very cold) with challenging access to communication/data systems from a ship at sea for up to 90 days.

Looking Ahead

How to create a Cyber Infrastructure design that promotes open and broad access to information and data while at the same time ensuring robust Cyber Security and reliability.

Mike Prince
ARV Project Manager
NSF/GEO/OPP

https://future.usap.gov/arv/

CI Accomplishment

Re-architect and implement IceCube's real-time alert directional reconstruction pipeline

CI Challenges or Opportunities

Flat funding with increasing science scope
Recruiting

Looking Ahead

Sharing resources/expertise
Cross-pollination between CISE and MF community

Benedikt Riedel
IceCube Computing Manager & Technical Coordinator
A3D3 Heterogeneous Systems Lead
UW-Madison

icecube.wisc.edu
a3d3.ai

CI Accomplishment

Complete ground-up redesign of the low-latency data distribution system for LIGO to enable real-time gravitational wave alerts.

CI Challenges or Opportunities

Reliable data management and access for our collaboration has been a very big challenge.

Looking Ahead

Managing complex CI for even "large" scientific projects is very difficult. How can we make this easier as a community?

Jameson Rollins
Control and Data Systems Architecture Lead
LIGO Laboratory
Caltech

https://ligo.caltech.edu

CI Accomplishment

The RCRV Program set ambitious goals: build 3 new Regional Class vessels and develop datapresence technologies to unlock observations for shoreside users. In 2023 we successfully introduced CORIOLIX, our datapresence implementation, to the Academic Research Fleet and now have 3 vessels online!

CI Challenges or Opportunities

Challenges: There have been many! Construction schedule setbacks coupled with emerging technologies (and obsolescence) make for tough procurement and lifecycle management of scientific CI. Introducing new technologies and workflow changes into an existing heterogeneously organized and operated facility has also been a challenge.

Looking Ahead

I'd like to echo what I've seen in a few of the other calling cards - a desire to build and support our cyber workforce! We already have recruitment and retention challenges given the unique lifestyle requirements for supporting Ocean Science, and we have work to do to develop (and fill) new positions with CI scope.

Chris Romsos
Datapresence Systems Engineer
Oregon State University
Regional Class Research Vessel

Oregon State University
College of Earth, Ocean, and Atmospheric Sciences

RCRV
Regional Class Research Vessels

https://ceoas.oregonstate.edu/regional-class-research-vessel-rcrv

CI Accomplishment

I am an active member of the FAIR Working Group and I have introduced the CI-Fellows to ORCID.

CI Challenges or Opportunities

How to introduce and integrate ORCID and other persistent identifiers into the CI-Compass infrastructure.

Looking Ahead

How to collaborate with Major Facilities when they are interested in adopting ORCID into their infrastructure and workflows.

Shawna Sadler
Engagement Manager,
Outreach & Partnerships
ORCID

https://orcid.org

CI Accomplishment

The Engagement and Performance Operations Center (EPOC) has worked with over 350 teams as part of our Roadside Assistance and Consulting Services, answering questions about end-to-end data movement across the full breadth of cyberinfrastructure.

CI Challenges or Opportunities

EPOC has worked extensively with many small and medium sized projects, but only a few of the major facilities (Arecibo, EHT, LHC, LCCF, NCAR, NRAO, VRO). This is a role we are expanding, especially with the ongoing need for larger scale data movement and our close ties to the LCCF.

Looking Ahead

We'd like to hear more about the problems that the major facilities are having with end-to-end workflows, looking at the full set of cyberinfrastructure problems they need to address to help their researchers be successful.

Jennifer M. Schopf
Director, EPOC
Director, Networking Partnerships
TACC / UT Austin

https://epoc/global

CI Accomplishment

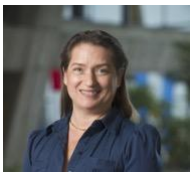
At the Compact Muon Solenoid, CMS, at the Large Hadron Collider in Switzerland, we organize the efforts of over 3 thousand scientists from all over the world. We develop a scientific software system of over 6 million lines of code and require large cyber infrastructure to do our discovery science.

CI Challenges or Opportunities

CMS's biggest challenge will come at the end of the decade when the experiment and accelerator will be upgraded. Those upgrades will result in an order of magnitude more data then the current CMS. That will mean having to process 10s of exabytes of data.

Looking Ahead

Looking ahead, CMS will need to invest significant effort in R&D to achieve its goals by the end of the decade. At the same time we are past the let all flowers bloom stage, and need to consolodate differently funded solutions into a coherent whole infrastructure across the globe. In other words more D then r.



Liz Sexton-Kennedy
CMS Software and Computing Co-Coordinator
Fermilab

OSG Council Chair

CI Accomplishment

In October, Trusted CI hosted our annual NSF Summit in Berkeley. The overall event was a major success. In particular, I'm proud of the student program organized by our team. In addition to providing students with mentors, we held a student poster session which drove some very productive and interesting discussion between CI professionals and students.

CI Challenges or Opportunities

Workforce development and diversity (demographic, experiential, and cognitive) remain a challenge. We are looking for ways to evolve our programs in future to help reduce the gaps in our community.

Looking Ahead

Speaking with one voice when communicating to funding agencies about community needs.

Kelli Shute
Executive Director
Trusted CI

Trustedci.org

CI Accomplishment

CyVerse managing cyberinfrastructure for NSF's new environmental synthesis center (ESIIL.org), and was also on the best short-version training manuscript at PEARC '23

Lead author on CyVerse: Cyberinfrastructure for Open Science, in press.

CI Challenges or Opportunities

Challenge: The proliferation of LLMs and GPTs; ensuring the equity and accessibility of commercial cloud & on-premises cloud; data management of private, sovereign, or sensitive data in this new age of AI. Opportunity: Open Source composable computing on federated environments.

Looking Ahead

Building collaborations and developing federated frameworks around cloud-native compute, data commons, and open science

Tyson Swetnam
Research Associate Professor
University of Arizona
CyVerse

CYVERSE

CyVerse.org

CI Accomplishment

Lead the CICompass Cloud Computing Working Group that focuses on helping Major Facilities Cloud migration effort. The working group just finished working on a report "NSF Major Facilities Cloud Use Cases and Considerations" that we are excited to make public.

CI Challenges or Opportunities

Majority of Major Facilities have been around for a long time. How do they update and migrate their CI while keeping the existing operations on-going.

Looking Ahead

More collaboration and exchange of ideas on how Major Facilities can benefit from leveraging academic and commercial cloud resources for their operations.

Karan Vahi
Senior Computer Scientist
CI Compass, USC

CI Compass
ci-compass.org

CI Accomplishment

This year we received 128 applications from over 50 schools for our 2024 CI Compass Fellowship Program for undergraduate students. We have an incoming class of 20 students from 17 schools across 11 states, who will be starting the Spring Program on January 23.

CI Challenges or Opportunities

I would love to expand our CICF Summer Program by getting more Major Facilities involved. I would also like to create more partnerships with HBCUs, TCUs, HSUs, and other MSIs.

Looking Ahead

I look forward to connecting with more Major Facilities. We would love MF input in the development and implementation of our CI Compass Fellowship Program.

Nicole Virdone
Project Manager and
Director of Outreach
CI Compass/ISI/USC

www.ci-compass.org

CI Accomplishment

At NOAA's National Centers for Environmental Information I lead the archive and access of NOAA's environmental data comprised of observations from the bottom of the ocean to the surface of the sun with over 60 PBs of data. Over the last few years the team has developed a knowledge graph based data management architecture to support the migration of the entire archive to the Cloud, enhance data discovery and access, and provide the foundation for AI/ML for NOAA and our partners.

CI Challenges or Opportunities

Using advancements in cyber technology there is a huge opportunity to improve the discovery, access, and interoperability of Federal government data to collectively address environmental, social, health, economic, and national security challenges. At the same time, there is a critical need to ensure the security and integrity of data. Advancing and improving collaborations across the government is critical to realize the full value of data, securely, and within limited budgets.

Looking Ahead

1) Opportunities and strategies to leverage development, co-development, and technically connect data repositories across government
2) How different agencies are handling data licensing
3) How to increase data citation and linkages of data to publications

Monica Youngman
Data Stewardship Division Chief

National Centers for
Environmental Information (NCEI)

National Oceanic and
Atmospheric Administration
(NOAA)

ncei.noaa.gov