# Broadening Student Engagement To Build the Next Generation of Cyberinfrastructure Professionals


Angela P. Murillo
Indiana University-Indianapolis, Luddy School of Informatics, Computing, and Engineering
apmurill@iu.edu

Don Brower
University of Notre Dame, Center for Research Computing
dbrower@nd.edu

Sarowar Hossain
Indiana University-Indianapolis, Luddy School of Informatics, Computing, and Engineering
sarhoss@iu.edu

Kerk Kee
Texas Tech University
kerk.kee@ttu.edu

Anirban Mandal
University of North Carolina, RENCI
anirban@renci.org

Jarek Nabrzyski
University of Notre Dame, Center for Research Computing
naber@nd.edu

Erik Scott
University of North Carolina, RENCI
escott@renci.org

Nicole Virdone
University of Southern California, Information Science Institutes
virdone@isi.edu

Rodney Ewing
Indiana University-Indianapolis, Luddy School of Informatics, Computing, and Engineering
raewing@iu.edu

Ewa Deelman
University of Southern California, Information Science Institutes
deelman@isi.edu



## ABSTRACT

The CI Compass Fellowship Program (CICF) was developed to broaden undergraduate student participation in cyberinfrastructure (CI) research, development, and operations. CICF is a distinctive program for undergraduate students pursuing studies in computer science, information science, data science, and other related fields. During year one of the program, CICF had six students participate from two institutions. During year 2 of the program, CICF had fourteen students participate from nine institutions. This poster provides details of the CICF program development and summarizes the impact of the first two years.



## CCS CONCEPTS

• **Applied Computing**; • **Education**; • **Information Systems**; • **Data Management Systems**;

## KEYWORDS

Student fellowships, workforce development, cyberinfrastructure, Major Facilities




*PEARC '23, July 23–27, 2023, Portland, OR, USA*



ACM ISBN 978-1-4503-9985-2/23/07.
https://doi.org/10.1145/3569951.3597567

**ACM Reference Format:**
Angela P. Murillo, Don Brower, Sarowar Hossain, Kerk Kee, Anirban Mandal, Jarek Nabrzyski, Erik Scott, Nicole Virdone, Rodney Ewing, and Ewa Deelman. 2023. Broadening Student Engagement To Build the Next Generation of Cyberinfrastructure Professionals. In *Practice and Experience in Advanced Research Computing (PEARC '23), July 23–27, 2023, Portland, OR, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3569951.3597567


## 1 INTRODUCTION

The CI Compass Fellowship program (CICF) was created by CI Compass to increase student participation in CI research, development, deployment, and operations, and assist with building the next generation of cyberinfrastructure professionals. CI Compass [1] is a National Science Foundation (NSF) Cyberinfrastructure Center of Excellence, which provides expertise and active support to cyberinfrastructure practitioners at NSF Major Facilities (MFs) to accelerate the data lifecycle and ensure the integrity and effectiveness of the cyberinfrastructure (CI) upon which research and discovery depend. The vision for CI Compass is to support and enhance the national CI ecosystem that includes people, practical knowledge, and processes to facilitate knowledge sharing and discovery across the MFs [1]. MFs represent the largest-scale research and engineering facilities in which NSF invests. These facilities often utilize shared-use infrastructure, instrumentation, and equipment that captures, processes, archives, and disseminates data. They are intended to serve a broad community of scientists, researchers, and educators, and are essential for scientific innovation [2]. To manage their complex and distributed resources, data, and processes, MFs develop, deploy, and operate sophisticated CI. This poster details CICF's purpose and structure, its relevance in fostering student development in the

**Table 1: CICF Spring Training Program Curriculum (Spring 2023)**

| Week | Technical Training Program | Research Training Program |
|---|---|---|
| Week One | Orientation, Linux/Unix Shell, Terminal | Cyberinfrastructure, Major Facilities, and the Data Lifecycle. Guest Speaker from the Natural Hazards Engineering Research Infrastructure |
| Week Two | Introduction to Python | Research Data Management, Guest Speaker from ORCID |
| Week Three | Python Programming, Jupyter Notebooks, Python Data Analysis Packages | Research Computing, Guest Speaker from the Texas Advanced Computing Center |
| Week Four | Best Practices in Software Development, Version Control, GitHub, Pytest | Research and Data Ethics |
| Week Five | Best Practices in Software Development, Containers, Docker | Guest Speaker from the National Center for Atmospheric Research |
| Week Six | Cloud Computing, Part 1 | Major Facilities and the Data Lifecycle, Part 1 |
| Week Seven | Cloud Computing, Part 2 | Major Facilities and the Data Lifecycle, Part 2 |
| Week Eight | Chameleon Cloud | FAIR Data |
| Week Nine | Spring Break | Spring Break |
| Week Ten | Data Workflows, Pegasus | Professional Skills, Networking |
| Week Eleven | Guest Speaker from the National High Magnetic Field Laboratory | Group Presentations, Day 1 |
| Week Twelve | Machine Learning/AI | Group Presentations, Day 2 |

context of CI and MFs, and summarizes the impact of the first two years of the fellowship program.

## 2 CICF PURPOSE

As outlined in the Missing Millions Report [3], funded by the Office of Advanced Cyberinfrastructure at the NSF, there is a significant challenge with CI workforce development and a lag in representation in the CI workforce. The report calls for broader engagement in CI projects, including expanded investments in apprenticeships, internships, and training grants to provide students with opportunities to participate in CI and CI-dependent research efforts. One way to increase representation in CI-related fields is to engage students through outreach and inclusion efforts and create potential career paths.

CICF was created to broaden student participation in the CI fields by providing undergraduate students with a way to learn about and engage with the CI community. The program provides CICF undergraduate student fellows the opportunity to:

- Learn about CI development and MFs.
- Develop CI-related skill sets important to the work of MFs.
- Engage with CI Compass and MF personnel through a virtual training and research program.

Through student involvement in CICF, CI Compass can create a pipeline for students to enter into the fields related to CI. The goal is to help students learn about MFs, scientific computing, and research computing and get them excited about potential opportunities that they likely do not learn about in traditional computer science, information science, data science, and related undergraduate programs.

## 3 CICF PROGRAM STRUCTURE

Structurally, CICF consists of a virtual Spring Training Program and an optional/invited Summer Hands-On Program. The virtual Spring Training Program has two components, 1) a technical training program and 2) a research training program. Table 1 provides an outline of the twelve-week Spring 2023 curriculum.

The *technical portion* of the spring program provides student fellows with experience in technical skills relevant to CI. Students learn and gain experience with basic software development, programming for scientists, systems, machine learning, and artificial intelligence relevant to CI and MFs.

The *research portion* of the spring program helps students understand the importance and context of MFs and the related data and CI. Student fellows learn about the data lifecycle of specific MFs and engage with guest speakers from MFs and the greater CI community. In groups, student fellows research a specific MF to learn about the science mission, CI, and data lifecycle at that MF. At the end of the spring program, student fellows present their research findings to CI Compass and MF colleagues.

The technical training program runs parallel to the research training program so that student fellows can gain hands-on technical experience and understand cyberinfrastructure implementation at MFs. Both the technical and research training sessions had guest speakers from various MFs, including NHERI [4], NCAR [5], MagLab [6], IceCube [7], and NEON [8]. Additionally, we had guest speakers from ORCID [9], TACC [10], and Pegasus [11] to provide students the opportunity to learn about and interact with CI professionals outside of MFs and see the connections between the greater CI community, scientific computing, and MFs.

Along with the Spring Technical and Research Programs, CICF student fellows may apply for an optional/invited Summer Program. The selected CICF student fellows participate in a hands-on,

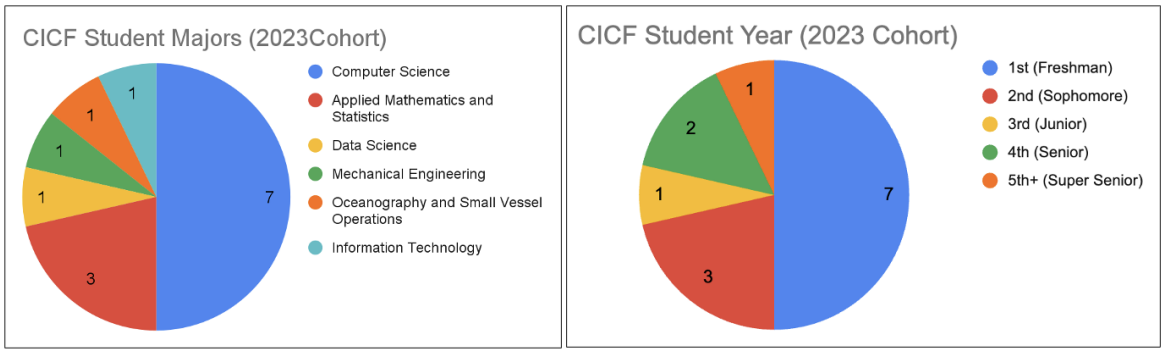


**Figure 1: (a) 2023 CICF Student Majors, (b) 2023 CICF Student Year**

project-based learning experience at either an MF or a CI Compass institution. During the 2022 Summer Program, three CICF student fellows gained hands-on, real-world experience at the University of Southern California. The student fellows worked with CI tools, including the Chameleon Cloud testbed [12], the Pegasus Workflow Management System [11], and HTCondor [13], to test and classify lake zooplankton to learn about the health of the ecosystem and the impacts of environmental changes. The student fellows wrote a short conference paper titled "Application of Edge-to-Cloud Methods Toward Deep Learning" [14] for IEEE eScience and presented the related poster [15] at the conference in Salt Lake City, Utah, in October 2022. During Summer 2023, CI Compass is collaborating with three MFs to place student fellows at MFs for hands-on projects. We have plans for both in-person and virtual participation for six of our 2023 student fellows at NCAR [5], NEON [8], and the Ocean Observatories Initiative (OOI) [16].

## 4 CICF PROGRAM IMPACT AND FUTURE PLANS

In its pilot phase, the 2022 CICF program had six student fellows participate from two institutions (the University of Notre Dame and the University of Southern California). The 2023 CICF program expanded to 14 students from nine institutions, including Indiana University, Louisiana State University, University of Iowa, University of Alabama, and Arizona State. The undergraduate student fellows came from six majors (see Figure 1a) and spanned from freshmen through 5th year/super seniors (see Figure 1b). Additionally, the 2023 cohort had eight female and six male students, three of which were first-generation students.

As described previously, student fellows worked in groups throughout the Spring program to research specific MFs. These end-of-program presentations demonstrate the knowledge gained throughout the program. The student fellows review both websites and published documents, as well as meet with MF staff to learn about the science mission, data lifecycles, and CI of the MF they choose to focus on for their research. During the Spring 2023 program, student fellows researched TACC/LCCF [10], MagLab [6], OOI [16], and the Cornell High Energy Synchrotron Source (CHESS) [17]. Their video presentations are publicly available on the CI Compass YouTube channel [18].

For the Spring 2023 program, we created a survey to compare improvement in student knowledge and interest at the beginning and at the end of the Spring Program. Through this pre/post-test survey, student fellows indicated an increased knowledge and interest in MFs, scientific computing, and CI. For example, one student described how much they learned about MFs during the program and how their interest has grown:

> "I knew nearly nothing about major facilities coming into this program. I left knowing so much and wanting to continue researching. I didn't think I was interested in scientific computing until this program" (F1)

Another student indicated their eagerness to continue exploring potential opportunities related to CI.

> "As a result of my participation in the program, I am now more eager than ever to explore opportunities in scientific computing and cyberinfrastructure." (F2)

Another student indicated their interest in research and pursuing a graduate program has significantly increased.

> "Since completing the program, my interest in pursuing undergraduate research or a graduate degree has grown substantially." (F3)

The survey results demonstrate that introducing students to MFs and CIs through CICF could broaden participation in the CI and MF workforce or, at the least, make students aware of the possibility of these career options.

CICF will continue to enhance the program by reviewing and updating the Spring curriculum, creating additional collaborations with MF colleagues for summer programs, and cultivating partnerships with other CI professionals. Next year, we plan to incorporate faculty mentors, student peer mentors, and continual outreach to universities and colleges for student recruitment to broaden the program's reach.

## ACKNOWLEDGMENTS

This project is supported by the National Science Foundation Office of Advanced Cyberinfrastructure in the Directorate for Computer Information Science under Grant #2127548.

## REFERENCES


[1] CI Compass. 2023. CI Compass. Retrieved April 17, 2023 from https://ci-compass.org/

[2] National Science Foundation. National Science Foundation Research Infrastructure Guide (NSF 21-107). Retrieved from https://www.nsf.gov/pubs/2021/nsf21107/nsf21107.pdf

[3] Alan Blatecky, Damien Clarke, Joel Cutcher-Gershenfeld, Deborah Dent, Rebecca Hipp, Ana Hunsinger, Al Kuslikis, and Lauren Michael. (2021). The missing millions: Democratizing computation and data to bridge digital divides and increase access to science for underrepresented communities. National Science Foundation. https://www.rti.org/publication/missing-millions

[4] The Natural Hazards Research Infrastructure (NHERI). 2023. DesignSafe | DesignSafe. DesignSafe | DesignSafe. Retrieved April 18, 2023 from https://www.designsafe-ci.org/

[5] National Center for Atmospheric Research. 2023. National Center for Atmospheric Research. Retrieved April 18, 2023 from https://ncar.ucar.edu/

[6] National High Magnetic Field Laboratory. National MagLab - MagLab. Retrieved April 18, 2023 from https://nationalmaglab.org/

[7] IceCube - The University of Wisconsin-Madison. 2023. IceCube Neutrino Observatory. IceCube. Retrieved April 18, 2023 from https://icecube.wisc.edu/

[8] National Ecological Observatory Network (NEON). 2023. NEON | Open Data to Understand our Ecosystems. National Ecological Observatory Network (NEON). Retrieved April 18, 2023 from https://www.neonscience.org/

[9] ORCID. 2023. ORCID - Connecting research and researchers. Retrieved April 18, 2023 from https://orcid.org/

[10] Texas Advanced Computing Center, The University of Texas at Austin. 2023. Texas Advanced Computing Center - Powering discoveries that change the world. Texas Advanced Computing Center. Retrieved April 18, 2023 from https://www.tacc.utexas.edu/

[11] Pegasus WMS. 2023. Pegasus. Pegasus WMS. Retrieved April 17, 2023 from https://pegasus.isi.edu/

[12] Chameleon Cloud. 2023. Chameleon. Retrieved April 17, 2023 from https://www.chameleoncloud.org/

[13] HTCondor. HTCondor Software Suite. Retrieved April 17, 2023 from https://htcondor.org/

[14] Khushi Choudhary, Nona Nersisyan, Edward Lin, Shobana Chandrasekaran, Rajiv Mayani, Loic Pottier, Angela P. Murillo, Nicole K. Virdone, Kerk Kee, and Ewa Deelman. 2022. Application of Edge-to-Cloud Methods Toward Deep Learning. In 2022 IEEE 18th International Conference on e-Science (e-Science), IEEE, Salt Lake City, UT, USA, 415-416. DOI:https://doi.org/10.1109/eScience55777.2022.00065

[15] Choudhary, Khushi, Nersisyan, Nona, Lin, Edward, Chandrasekaran, Shobana, Mayani, Rajiv, Pottier, Lo√Øc, Murillo, Angela P., Virdone, Nicole K., Kee, Kerk, and Deelman, Ewa. 2022. Application of Edge-to-Cloud Methods Towards Deep Learning. Zenodo. Retrieved April 17, 2023 from https://zenodo.org/record/7140858

[16] Ocean Observatories Initiative. 2023. Ocean Observatories Initiative. Ocean Observatories Initiative. Retrieved April 18, 2023 from https://oceanobservatories.org/

[17] Cornell Research. 2023. Cornell High Energy Synchrotron Source (CHESS). Cornell High Energy Synchrotron Source (CHESS). Retrieved April 18, 2023 from https://research.cornell.edu/content/cornell-high-energy-synchrotron-source-chess-0

[18] Spring 2023 CI Compass Fellowship Final Student Presentations. 2023. CI Compass. Retrieved April 18, 2023 from https://www.youtube.com/watch?v$=$LFXGX--_Ewo