Building a Cybertraining program for Climate Scientist in the Pacific to integrate Cyberinfrastructure and Open Science

SEAN B. CLEVELAND, University of Hawai'i – System, USA
SHIVANI TANAKA, University of Hawai'i – System, USA
MARIA DUMANLANG, University of Hawai'i – System, USA
ALEXANDER J. STOKES, University of Hawai'i at Manoa, USA
JASON LEIGH, University of Hawai'i – Manoa, USA
THOMAS W. GIAMBELLUCA, University of Hawai'i at Manoa, USA
PHILIP M. JOHNSON, University of Hawai'i – Manoa, USA
HELEN TURNER, Chaminade University of Honolulu, USA
GWEN A. JACOBS, University of Hawai'i – System, USA

Abstract:

The Cyberinfrastructure Training and Capacity Building in Climate and Environmental Sciences (CI-TRACS) program represents a pioneering initiative aimed at enhancing cyberinfrastructure proficiency within Hawaii's academic community. This paper outlines the program's comprehensive strategy, which integrates curriculum development, hands-on workshops, and professional growth opportunities to cultivate a robust foundation in CI practices. The initiative's core objective is to elevate CI literacy, promote cross-disciplinary cooperation, and endorse the principles of open science.

Significant contributions from the CI-TRACS program include a suite of educational materials and resources tailored for integration into higher education syllabi. Collaboration with the Hawaii Data Science Institute has been instrumental in nurturing a burgeoning network of data science professionals. The CI-TRACS program is instrumental in realizing the shared vision of equipping Hawaii's emerging workforce with the sophisticated CI skills necessary to navigate and excel in the evolving landscape of climate and environmental sciences.

CCS Concepts: • **Social and professional topics → Computing education programs**.

Additional Key Words and Phrases: Professionals, contributors, undergraduates, graduate students, climate science, research, CI skills.

ACM Reference Format:
Sean B. Cleveland, Shivani Tanaka, Maria Dumanlang, Alexander J. Stokes, Jason Leigh, Thomas W. Giambelluca, Philip M. Johnson, Helen Turner, and Gwen A. Jacobs. 2024. Building a Cybertraining program for Climate Scientist in the Pacific to integrate Cyberinfrastructure and Open Science. In *Practice and Experience in Advanced Research Computing (PEARC '24), July 21–25, 2024, Providence, RI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3626203.3670617

1 INTRODUCTION

The study of climate science in Hawaii is both a priority given regional impacts of climate change and an opportunity by virtue of the unique terrestrial and marine environment. Hawaii and other Pacific Islands are at the forefront of

Manuscript submitted to ACM

the negative impacts of climate change. The CI-TRACS program (CyberInsfrastructure TRaining to Advance Climate Science) supports cyberinfrastructure (CI) training for environmental science research, education, and practice in the Hawaii-Pacific region. This program has created workshops and curriculum modules for undergraduate and graduate students that increase CI skills across the environmental science domain. By providing the next generation of environmental scientists and research practitioners with advanced CI skills, the potential for transforming science in the region across academic research institutions, public agencies, and community stakeholders is significant.

The three-year project that started in 2021 has built a multi-level capacity enhancement program for CI training and skills development that serves CI users, contributors, and professionals in the environmental and climate science communities. 1.) The training materials in the workshops and curriculum modules can be incorporated in undergraduate and graduate courses in order to promote CI skills development and awareness. 2.) The project has increased the pool of CI professionals and facilitated knowledge transfer from existing CI professionals to the environmental and climate science communities. 3.) The project aims to advance domain-specific CI skills development to position students as CI contributors in their future professional careers. The project has delivered important impacts to the environmental and climate science communities in the Pacific Island region as well as to others in and beyond the region. 4.) The project is conducted in the classroom location of the Hawaii Data Science Institute, which is positioned as an ideal platform for longitudinal engagement and interdisciplinary collaboration, supporting open science.

In this paper, we present the implementation and evolution of the CI-TRACS program and lessons learned.

2 BACKGROUND

Climate change research is both interdisciplinary and very data intensive spanning atmospheric, hydrology, ecology and ocean sciences. There is a significant opportunity to improve research productivity in climate science at University of Hawaii (UH) and in the Pacific Islands through the adoption of advanced cyberinfrastructure skills, technologies and access to national resources. Prior to this project there was no cohesive educational program at UH or Chaminade that provides training in basic and advanced CI skills necessary for climate research. Typically students acquire a limited subset of necessary skills in an ad hoc fashion from their advisors and receive no formal training or exposure beyond the techniques used in their labs. Students don't know what they don't know or who to ask for help. Accessing CI resources beyond a laptop/workstation is a daunting barrier to many students, especially if their advisors lack these skills or lack experience in this area. There is an over abundance of CI tools, science gateways and data resources available for climate research and choosing the appropriate tools is challenging and confusing. Many software packages function like black boxes to students and their limited understanding of the tools can result in poor science and lack of reproducibility. Most importantly, climate science research requires integration of diverse data types, models and simulation results. This integration requires advanced, sophisticated CI approaches to produce reproducible results and accurate predictions. CI-TRACS is designed to fill this training gap for climate scientists.

Since its inception in 2018, the Hawaii Data Science Institute (HIDSI) continues to be a hub for supporting collaboration, research opportunities and workforce development across the UH community and partner institutions. The Data Science Fellows Program was developed in the HIDSI with the intent to broaden the adoption of CI and Data Science (DS) skills, support the creation of CI and DS curriculum resources and provide opportunities for professional development.

3 CI-TRACS PROGRAM

The modern scientific research enterprise and data life-cycle is comprised of phases including project conceptualization/brainstorming, data collection & cleaning, data analysis & visualization, knowledge crystallization, and knowledge

Demographics of graduate student participants				
Classification	**Award Year 1**	**Award Year 2**	**Award Year 3**	**Total**
Total participants	10	14	11	35
Male	6	8	4	18
Female	3	1	5	9
URM/ NHPI	0	0	1	1
1st Gen college student	1	1	1	3
Economically Depressed Area	4	0	2	6
Declined to report	1	5	2	8

Table 1: Demographics of undergraduate student participants in the CI-TRACS program. Totals include individuals who declined to report demographics.
URM: Underrepresented minority including individuals who identified as Black, African American, Hispanic, Native American and Alaskan Native.
NHPI: Native Hawaiian or Pacific Islander.

dissemination & presentation, all of which can be enhanced through CI [Leigh2019]. The CI-TRACS program provides learning opportunities to master CI skills for each phase of the research enterprise through a series of targeted workshops designed and taught to our stakeholders by CI Fellows, a summer immersion program for undergraduates and the development of a repository of shared curriculum training materials.

3.1 CI Fellows

The academic year program was designed to meet two needs: 1) focused immersive workshops in CI skills for climate science researchers and 2) broad training for CI Fellows who participate as trainers in workshops. CI Fellows are recruited from departments (climate science [CLS], computer science [CS] and data science [DS]) with the goal of achieving a balance between CS and DS students and those in atmospheric, geo and ocean sciences. Professional development activities are programmed jointly with other graduate students in the HIDSI Data Science Fellows Program [HI-DSIFellows]. Activities include portfolio development in GitHub; speakers/discussion on STEM inclusion and bias, speaker/discussion on research ethics, data sovereignty and mentoring/career development panels.

Over the last three years, the program has recruited a diverse cohort of students in good academic standing with disciplines in computer science and natural sciences. A total of thirty-six graduate students representing 11 departments, including Atmospheric Science, Economics, Electrical Engineering, Finance, Information and Computer Science, Marine Biology, Mathematics, Natural Resources and Environmental Management, Tropical Conservation Biology and Urban and Regional Planning, participated in the year-long training program. Twelve students, four each year, were directly funded by the CI-TRACs program. Other funding sources were leveraged to support other participating students. Participant demographics are represented in Table 1. Twenty-seven participants chose to report demographics- 25% identified as female and 8% were first generation college students. As part of the program, graduate students, in collaboration with workshop instructors developed and co-presented 10 immersive workshops each year. Workshops were structured as hands-on modules to enhance the learning experience and maximize participant engagement. The workshop repository is publicly available on the Morea Framework: https://change-hi.github.io/archive-2022-2023/modules/. Participants presented their learning experience as a workshop series to the UH community and general public. Video recordings of workshops are publicly available on the Hawaii Data Science Institute YouTube channel (https://www.youtube.com/@hawaiidatascienceinstitute6736/videos).

Fellows applied cyberinfrastructure skills acquired through the program to their thesis research and presented these applications at 6 research symposia, at the end of the fall and spring semesters each year. The year-long program includes a mentorship experience where data science fellows apply acquired skills and guide undergraduate students through the life cycle of a data science project during an undergraduate summer program (CITRUS).

HIDSI and Hawaii EPSCoR Change-HI project's training efforts and resources were leveraged to build a community of practice. Data Science Fellows met weekly in the classroom hub of the Hawaii Data Science Institute to share updates on research and present workshops. In year 3, a coffee hour for students was introduced as an informal platform for students to discuss research experiences and projects. Workshops to the public included the applications of low-code and no-code tools (Tableau, SAGE3) to design web apps and conduct data analysis. Faculty all-hands meetings were intentionally combined with student research presentations to increase cross-collaboration, foster inclusiveness and reduce barriers between faculty and fellows. Professional development opportunities included research symposia, conferences and seminars featuring invited speakers from Chaminade University and partner organizations.

3.2 Workshops and Modules

During the academic year, professional UH cyberinfrastructure personnel, HIDSI faculty and CI Fellows design and deliver a set of hands-on workshops focused on a set of topics designed to span the research life-cycle of climate science projects, starting with problem conceptualization to data dissemination and archiving. These workshops target the sets of basic and advanced skills necessary for climate scientists working across three broad areas: atmospheric science, hydrology and ecology. A total of 36 workshops were developed during years 1-3 of the program. Demographics of attendees at each workshop are represented in Table 2. 89

The first year of workshops built on efforts from the CI community such as The Carpentries [2]. The workshop format was adopted from The Carpentries style and some of the initial content for some of the skills such as command line and git were leveraged and integrated (https://github.com/CI-TRACS). However, based on discussions after the first year of the program it was identified that some of this organization of an overall curriculum was lacking and the workshops presented as more standalone content, rather than a curriculum. Therefore, in year 2 workshops were restructured as modules and reorganized to enhance the learning experience and maximize participant engagement (https://change-hi.github.io/modules/). Graduate student fellows led the development and design of an immersive workshop series utilizing CI skills. In collaboration with workshop mentors, students met weekly and restructured workshop module content on the Morea Framework [1]. Participants presented their learning experience as a workshop series to the UH community and general public.

3.3 Undergraduate Summer Immersion Program

The CITRUS program (CyberInfrastructure TRaining for Undergraduates in Summer) program provides a one-month REU-style immersion combining development of key skills and abilities in CI with a climate-focused research project applying CI skills. 187 students applied to the undergraduate program during the 3 years of this award. A total of 38 undergraduate students, of whom 36% were first generation college students, 39% female and 21% from economically depressed areas participated in the program (Table 3). The program was held virtually to enable participation from students located on neighboring islands and the mainland. Each undergraduate participant was paired with a CI Fellow. Research themes centered around the impacts of climate change on Hawaii's environment, health and economy. Students focused on project ideation, moving, cleaning, organizing data, analyzing data sets and creating visualizations. Mornings were used for self-paced data skills training modules or CI-TRACS module workshops taught by CI Fellows

CI-TRACS Cumulative Workshops Metrics				
Classification	**Year 1**	**Year 2**	**Year 3**	**Total**
Workshops	11	13	12	36
Attendees	145	237	196	578
Male	74	121	76	271
Female	61	81	103	245
Other	10	88	91	189
URM	18	32	39	89

Table 2: A snapshot of workshop metrics and demographics of attendees for all three years of the program. The workshop totals do not necessarily represent total individuals but is an aggregation of attendance at each workshop.

Demographics of undergraduate student participants				
Classification	**Award Year 1**	**Award Year 2**	**Award Year 3**	**Total**
Total participants	12	10	16	38
Male	4	7	7	18
Female	5	3	7	15
URM/ NHPI	3	1	5	9
1st Gen college student	5	3	6	14
Economically Depressed Area	4	1	3	8

Table 3: Demographics of undergraduate student participants in the CI-TRACS program. Totals include individuals who declined to report demographics.
URM: Underrepresented minority including individuals who identified as Black, African American, Hispanic, Native American and Alaskan Native.
NHPI: Native Hawaiian or Pacific Islander.

and afternoons for meetings with near peer mentors. At the end of the program, students presented at a research showcase.

3.4 Assessments and Feedback

Workshops and overall program surveys were conducted to assess the curriculum and general student experience. At the end of each workshop, a survey was conducted to evaluate the presentation, pedagogy and content of the workshops in order to provide feedback for the CI fellows and mentors as well as inform the overall project as to what could be improved in the workshop experience. All participants and workshop instructors were highly encouraged to respond to surveys and a dedicated time was incorporated to each workshop to facilitate this activity. Surveys were conducted through Google forms and included open-ended questions assessing content recall, self-reported usefulness of content, applicability to current research, potential for future applications and collaborations, recommendations for changes as well as additional topics attendees would like featured. Undergraduate and graduate students found this to be a positive experience overall. Graduate students appreciated being able to work with peers and found the mentorship experience to be rewarding. Several undergraduates in the CITRUS program mentioned that they had acquired new cyberinfrastructure skills or improved existing knowledge of data science applications and available resources in surveys

administered at the end of the summer program. Additionally, a survey to the University community was disseminated at the end of year 2, inquiring about additional workshop topics for CLS and CI that could be developed and added. Based on feedback, 4 new workshops around the Hawaii Climate Data Portal, Experimental Time Series Analysis in R, Software Code Quality and Creating Professional Portfolios were added. Two new seminars on Data Inequities and CARE Data Principles were also added as a result of this survey.

4 LESSONS LEARNED

In order to better align with campus wide and departmental recruitment for TA positions, student recruitment for the graduate program would need to begin in December of the previous year. Based on feedback from previous cohorts, in year 3, a workshop about best practices on running a workshop was introduced. Guidelines on how to conduct workshops and support participant engagement were documented and shared with workshop mentors. Workshop assessments indicated that workshop content was useful though attendees mentioned that they would like better support for a community of practice. Future efforts will focus on creating opportunities for engagement and supporting a community of practice, including continuing a monthly coffee hour gathering for graduate and undergraduate students to engage in discussion. Plans to scale this program statewide include dissemination of workshop information and events conducted through the Hawaii Data Science Institute to other UH campuses and community college students served by the UH System. Overall feedback from students indicated that student led workshop topics offered more ownership for the CI Fellows and supported participant engagement.

5 IMPACTS

Data science fellows participating in the program have opportunities to leverage research opportunities through the Hawaii Data Science Institute and EPSCoR Change(HI) project. Graduate students have the opportunity to present their work to a broader research community at national and international conferences. Data science fellow Gerardo Rivera Tello published a Nature paper (doi.org/10.1038/s41598-023-45739-3) during his participation in the program. Beyond the term of the program, alumni are encouraged to join the Hawaii Data Science Institute's LinkedIn and share their career achievements, including research publications and awards. Two program alumni were recruited by UH and are now research software engineers. Several alumni currently collaborate with UH researchers on ongoing research projects and remain connected by sharing their experiences with newer cohorts. Other alumni continue to attend workshops, seminars and symposia. One of the CITRUS student participants is entering graduate school this year. The program serves as a platform for interdisciplinary collaboration by leveraging the expertise of faculty from different departments and continuing to recruit participants with various academic disciplines.

6 CONCLUSION

The CI-TRACS program has significantly advanced cyberinfrastructure (CI) capabilities within Hawaii's environmental and climate science sectors. By implementing a comprehensive strategy that includes the development of specialized cyberinfrastructure and data science curricula aimed at climate scientists, the execution of immersive workshops, and professional development opportunities, the initiative has effectively enhanced CI proficiency at UH and across the state. The collaborative efforts with the Hawaii Data Science Institute have been instrumental in cultivating a robust network of cyberinfrastructure and data science professionals, thereby enriching the research community. The positive outcomes of the program are reflected in the affirmative feedback from participants and the observable benefits in the educational and professional trajectories of students and alumni. The CI-TRACS program is thus a cornerstone in the

ongoing endeavor to equip Hawaii's emerging workforce with the sophisticated CI and data science skills necessary to navigate and excel in the increasingly data-driven global landscape.

ACKNOWLEDGMENTS

This work was made possible by grant funding from the following: NSF OAC-2118222 CyberTraining: Implementation: Medium: Cyberinfrastructure Training to Advance Environmental Science, NSF OIA-2149133 RII Track-1: Change Hawaii: Harnessing the Data Revolution for Island Resilience, NSF CSE-1920304 MRI Acquisition: High Performance Computing Cluster for Data Intensive Research, NSF OAC-2201428 CC* Compute: Koa - A High Performance and Flexible Research Computing Resource and NSF OAC-2232862 CC* Data Storage: KoaStore: A High Performance and Flexible Research Storage Resource, NSF 2221448 Scholarship project supporting undergraduate and graduate students at the interface between biology and data science, NSF 2125483 BSCER Investigator Initiated Development program. Justice-orientated Approaches for Data science Education (JADE) and NSF 2217242 Collaborative Research: NSF INCLUDES Alliance: Alliance Supporting Pacific Impact through Computational Excellence (ALL-SPICE).

REFERENCES

[1] Philip Johnson. [n. d.]. Morea web site. https://morea-framework.github.io/.
[2] Greg Wilson. [n. d.]. Carpentries web site. http://carpentries.org.